BEYOND THE RECEPTIVE FIELD: AN ANALYSIS OF NATURAL SCENES AND A GEOMETRIC INTERPRETATION OF EFFICIENT CODING STRATEGIES BY THE MAMMALIAN VISUAL SYSTEM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Kedarnath Padmakar Vilankar

August 2017

BEYOND THE RECEPTIVE FIELD: AN ANALYSIS OF NATURAL SCENES AND A GEOMETRIC INTERPRETATION OF EFFICIENT CODING STRATEGIES BY THE MAMMALIAN VISUAL SYSTEM

Kedarnath Padmakar Vilankar, Ph.D.

Cornell University 2017

In biological and artificial neural networks the response properties of a visual neuron are often described in terms of a two-dimensional response map called the receptive field. This receptive field is intended to capture the basic behavior of a neuron and predict how that neuron will respond to a novel stimulus. However, the receptive field provides a good description of the neuron's behavior only if the neurons in the network are linear. Neurons in an organism are in fact highly nonlinear, which means their responses are not completely described by their receptive fields. A number of studies have attempted to explain the properties of these neurons in terms of an efficient representation of natural scenes. In this thesis I will demonstrate the hidden computations and interactions a network of neurons performs which are not described by their receptive field.

In the first study (Chapter 2), I address an aspect of natural scenes that is rarely considered in discussions of efficient coding. This study explores how the structural properties of an edge relate to the cause of the edge. I will show that neurons at the earliest stages of the visual system rather than just detecting edges (as depicted by their receptive fields) could potentially use these structural properties to identify the causes of an edge.

The next three studies (Chapters 3,4, and 5), I explore the non-linear re-

sponse of neurons. Most neurons in the visual pathway are nonlinear. To account for their behavior, we need an approach that goes beyond the classic receptive field. A variety of different approaches has attempted to explain this behavior. I present a geometric framework which attempts to provide a better description of the nonlinear response properties of neurons in the sparse coding network. I explore the geometric characterization of neurons in the efficient coding mechanisms like gain-control, a "fan equation" model for optimal sparsity, and a cascaded linear-nonlinear model. This geometric approach provides a deeper understanding of why sparse representations (including those of cortical visual neurons) give rise to nonlinear responses. The nonlinearities in artificial neurons are visualized and quantified in terms of the curvature of iso-response surfaces. I show that the magnitude of nonlinearities increases as the overcompleteness of the network increases, even though the linear receptive fields appears to be similar.

In the next study (Chapter 6), I explore and define two forms of selectivity based on the curvature of the iso-response surfaces. The first form is "classic selectivity", which is the stimulus that produces the optimum response from a neuron. The second form is "hyperselectivity" which is defined by the drop-off in response around the optimal stimulus due to the curvature of the iso-response surfaces. I show that the hyperselectivity is unrelated to the classic selectivity. For example, it is possible for a neuron to be narrowly tuned (hyperselective) to a broadband stimulus. Further, I show that hyperselectivity in a neurons response profile breaks the Gabor-Heisenberg limits.

Finally (Chapter 7), I show the effect of different learning rules, enforced by various cost functions used in the sparse coding network, on the response geometry of neurons. I demonstrate how different learning rules affect the interac-

tion between the neurons in three-dimensional networks and the implications these findings have for a better representation of natural scene data in higher dimensions of image state space.

BIOGRAPHICAL SKETCH

Kedarnath Vilankar was born in Agra, India. He completed his Bachelor of Engineering in Information Technology from University of Mumbai, India in 2007, then worked as a Software Engineer for Amdocs Pvt. Ltd for two years. He then completed Master of Science in Electrical engineering from Oklahoma State University, USA in 2012. Since 2012, he has been a graduate student in Cornell University, USA.

ACKNOWLEDGEMENTS

Many have supported me in my graduate school journey at Cornell University. Foremost, I would like to thank my advisor Professor David Field for his invaluable guidance, expert advice, and mentoring. I would not be where I am today without Professor David Field. I would also like to thank other members of my academic committee Professor James Cutting, Professor Shimon Edelman, and Professor Thomas Cleland for suggestions and feedback.

I owe many thanks to James Golden for all his support, guidance, and being a wonderful lab mate at Cornell University. I would like to thank Dr. Damon Chandler for introducing me and encouraging me to do research in vision science. Thanks to all the members of Perception, Action, and Cognition lab (Professor Khena Swallow, Jordon Delong, Catalina Iricinschi, kaitlin brunick, Ayse Candan, Kacie Armstrong) for suggestions and feedback on my published articles and talks.

Finally, I would like to thank my family and friends (Gayathri, Ashwini, Satya, and Ewa) for everything they have done for me.

I acknowledge the financial support from Google Faculty Research Award to David Field.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The mammalian visual system is an exquisite information-processing device. The visual system extracts information from the light that is reflected or emitted in the natural environment. Photons reflected from objects enter our eyes, where photoreceptors located in the retina absorb and convert the light energy into electrical signals. From here onwards the electrical signals are processed by a hierarchy of visual processing areas in the brain, where each is composed of millions of neurons. The knowledge acquired enables an organism to make important behavioral decisions for survival and reproduction. Vision has been a field of inquiry as far back as Aristotle, and exactly how the visual system processes information has been a question of great interest over the last century. We have come a long way and gained great insights into the mechanisms of information processing in the visual system. The work presented in this dissertation is an attempt to further that knowledge.

1.1 The physiological approach

Vision is the result of computation by many individual neurons. One of the first approaches to understanding what each neuron represents comes from the invention of a physiological technique called single-cell recording. In this technique, electrodes are placed near a neuron to measure the voltage fluctuations. This technique allows the experimenter to determine the stimulus (a 2D image pattern) that produces a response in the neuron. Kuffler (1953), used single-cell recording to determine the receptive fields (2D image patterns that produce a response) of retinal ganglion cells in the mammalian visual system. He found

that specific visual patterns at a particular location in the visual field either excites or inhibits the firing pattern of the neuron. He further found that the receptive field that produces a vigorous activity in retinal ganglion cell had center surround organization, that is either a bright center with a dark surround or a dark center with a bright surround stimulus. These cells were initially believed to be spot detectors.

Barlow (1953) found that some ganglion cells in the frog retina were responsive to a black disc moving back and forth in the visual field, and such stimuli produced a consistent jumping and snaping response behavior. Such findings made researchers believe that these neurons are "bug detectors" and part of the primitive form of recognition. Hubel and Wiesel (1959) used single-cell recordings to determine the receptive field of cells in the striate cortex of the cat. They found receptive fields to be elongated and oriented arrangements of dark and bright bars. Initially, these cells were considered to be edge and bar detectors. Hubel and Wiesel (1962) speculated that cells in the striate cortex sum the outputs of several aligned center-surround cells from lateral geniculate nucleus (LGN). These findings from single cell recordings gave rise to the general idea that each stage of the visual cortex in the ventral stream is encoding some visual feature and each successive stage sums up outputs of the previous stage to encode complex visual features. However, the work presented in this dissertation will argue that the neurons in the visual system are not simple feature detectors. The features estimated from the receptive fields are not the complete descriptions, and the neurons perform computations beyond those of the simple feature detection. It is widely accepted that the visual system has two streams of processing in the brain. Each stream consists of successive regions of the brain processing the visual information. The ventral stream (also referred to as "what

pathway") is involved in object recognition and the dorsal stream (also referred to as "where pathway") is involved in processing the object location. However, recent evidence suggests that the neural circuitries in the cortex are highly overlapping and each region in the brain is involved in multiple cognitive domains (Anderson, 2010; Anderson and Pessoa, 2011).

1.2 The computational approach

The physiological approaches yielded many insights into the processing of individual neurons. However, very little was understood about the information processing by networks of many neurons. With the increase in computational power in the 1970s, many vision scientists used computational approaches to investigate visual processing by simulating an array of interacting neurons. Marr (1982) proposed a new computational approach to understanding the biological visual system. Unlike physiologists who just looked at the features of the world that excited the neurons, Marr advocated a new structured framework to understand complex information processing systems. He argued that extracting relevant facts about the world is not enough, and that understanding how information gets represented internally is equally important. Marr (1982) summarized the study of any complex information processing system at three different levels: 1) the computation, 2) the representation and algorithm, and 3) the implementation. The computational level determines the abstract computational theory of the system. At this level, questions about the computational goal of the system are determined without considering how it is accomplished. At the representation and algorithm level, the algorithmic description of how the computations are executed is studied. Finally, the implementation level specifies

how the representation and the algorithm can be physically implemented. The work presented in this dissertation will focus on the computational goals and the algorithmic representations in the visual system.

1.3 Natural scene statistics

In the field of experimental psychology, use of naturalistic stimuli was emphasized by Brunswik (1947, 1952) and Gibson (1950). It was recognized that understanding the natural world is relevant to understanding the visual system that encodes it. However, the notion of efficient coding which takes advantage of the regularities and the statistics of the natural environment was first proposed by Attneave (1954) and Barlow (1961). They proposed the "efficient coding hypothesis" which states that the goal of any sensory system is to encode the statistics and the regularities of the world in which they evolved. Attneave (1954), applied techniques of information theory (Shannon, 1949) and argued that the natural visual signal is highly redundant. Attneave (1954) and Barlow (1972) argued that an efficient visual system must get rid of the redundancy present in the natural environment. Barlow (Barlow, 1953, 1961; Barlow et al., 1967; Barlow, 1972, 1979), stressed that it is essential to study the redundancies and the regularities of the natural world to understand the visual system better.

Following the efficient coding hypothesis of Barlow, many vision scientists in 1980s started using tools and techniques from digital image processing to study statistics of natural scene images. Although the image processing community used naturalistic images to develop the algorithms, but naturalistic images were not studied for the purpose of quantifying their statistical properties.

Field (1987) first used a set of digitized natural images to study their statistical properties. He was initially criticized for using the naturalistic stimuli, as the naturalistic stimuli were considered to be difficult to control. Before Field, most spatial vision experiments were performed using non-naturalistic stimuli such as gratings, spots, plaids, etc (e.g., Barlow et al. (1967); De Valois et al. (1978, 1982); Movshon et al. (1978a,b)). Soon after Field (1987), the natural scene statistics approach became popular and many insights were gained about the efficient encoding mechanism of the mammalian visual system (e.g., Atick (1992); Webster and Mollon (1997); Webster and Miyahara (1997); Lewicki and Olshausen (1999); Lewicki and Sejnowski (2000); Geisler et al. (2001); Murray (2013)). However, there still remains a debate over complex natural stimuli verses simple synthetic stimuli (Rust and Movshon, 2005).

Field (1987) applied Fourier analysis to six natural images to estimate the redundancy. He found that the amplitude of Fourier components is inversely proportional to the frequency. He observed that the natural image have most of the energy in the low frequencies as compared to the high frequencies, which implied that there are redundancies in the natural images in correlations between the intensities of nearby pixels. Field further analyzed multiple coding schemes and found that the coding properties of V1 simple cells are well-suited to represent the natural scene images. He demonstrated that simple cells modeled as a bank of Gabor functions produce a sparse response distribution, which implies that the Gabor code gets rid many of the higher-order redundancies of the pixel intensities. Kersten (1987) also showed that the natural images are highly redundant and are only a small proportion of all possible images. He performed a psychophysical experiment where human observers were asked to replace missing pixels in images. Using the correct guesses, he estimated that

the perceptual information content in a pixel was around 1.4 bit. Atick (1992) discussed design principles for sensory systems based on information theory and demonstrated that the models of the early visual system (retinal ganglion cells) based on these principles predict the physiological and psychophysical observations (Atick and Redlich, 1990). Since Field (1987), many have applied the tools of digital image processing to quantify the natural scene statistics. Studies of natural scene statistics have gained many insights into the efficient coding mechanisms of the visual system. In Chapter 2, I will present some of the natural scene statistics of the different categories of edges in the natural environment and will demonstrate that the early visual system can use these statistics to make a probabilistic decision about the categories of edges.

1.4 The statistics of edges in natural scenes

It is estimated that in mouse there are at least 30 or more types of retinal ganglion cells (Sanes and Masland, 2015). However, there is no clear taxonomy for these different types of cells, and they are not consistent across species. There are three basic groups of the ganglion cells: W-, X- and Y-ganglion cells in the cat. W-ganglion cells are the smallest of three and detect movement in the entire visual field. X-ganglion cells are medium in size, which are mainly responsible for color vision. Y-ganglion cells are the largest in size and respond to rapid eye movement or a change in light intensity. Based on projections and function of the ganglion cells there are five additional classes (Martin and Grünert, 2004). However, most visual modeling efforts only consider the one type of ganglion cell with on-off center surround organizations and one type of simple cells with elongated and oriented inhibitory and excitatory regions. The discovery of mul-

tiple types of ganglion cells begs the question if there are multiple computations being carried out in the initial stages of the visual system. In Chapter 2, I will relate how the early stage of the visual system can segregate different types of edges in natural scene images.

It is believed that one of the goals of the visual system is to provide an efficient representation of natural scenes by encoding the object boundaries and object layout. The early stages of the visual system identify these object boundaries by identifying the contrast changes in the visual scene. Hubel and Wiesel (1962), probed a number of cells in a cats striate cortex and found them to be responsive to stimuli with elongated dark and bright bars at specific orientations. Initially, these cells were believed to be edge and bar detectors. However, Marĉelja (1980) described these cells more appropriately with Gabor functions, which were Gaussian envelopes modulated by sinusoidal waves. The Gabor functions were later demonstrated to be a good description of the receptive fields of simple cells (e.g., Field and Tolhurst (1986); Jones and Palmer (1987)). The Gabor functions were observed to produce sparse representations, where only a few of the neurons responded to any natural stimuli, and most of the neurons responded only a bit or not at all (Field (1987)). This implies that the Gabor-like receptive fields provide an efficient representation of edges in natural scenes because it gets rid of the redundancies (such as higher-order correlations between pixels) and represent only the important changes in the images such as edges.

Abrupt changes in luminance or reflectance produce an edge. For example, an edge can be formed when one object occludes another object. These occluding edges define the object boundaries. These edges are important from an ob-

ject recognition perspective, as identifying these edges helps in segregating the object boundaries from the background. Other types of edges can arise from a reflectance change, a shadow, or an illumination boundary. It is quite possible that an edge could have multiple causes. However, it appears that the early stages of visual system do not produce any distinction in the representation of the causes of an edge. In chapter 2, I describe a study where we investigated the statistics of edges and estimated the probability of occurrence of different causes of edges in natural scenes. We calculated the local properties of edges such as contrast. The information like local contrast can be easily estimated from the response of the early stages of the visual system. Our goal was to determine if the local statistics of an edge could provide any information about the class of an edge at the early stages of the visual system.

1.5 Linear modeling of the visual system

As mentioned before, neurons of the striate cortex were initially thought of as edge detectors, but were later described by Gabor functions (Marĉelja, 1980). Marcelja drew many similarities between the Gabor functions and the oriented receptive fields of simple cells. Later, multiple studies showed that Gabor functions are accurate descriptions of simple cell receptive fields (e.g., Field and Tolhurst (1986); Jones and Palmer (1987)). Field (1987), demonstrated that Gabor-like functions are best suited to reduce the redundancies in natural scene images represenattions. Field showed that the natural environment is highly redundant and an efficient coding mechanism should get rid of these redundancies (following the argument by Barlow (1972)). Field found that Gabor-like codes gets rid of the higher-order redundancies (e.g., redundancy between multiple

pixels which form an edge) and convert them to first-order redundancy (i.e., a non-Gaussian and therefore predictable probability distribution). Further, he proposed that the Gabor representation produces a response distribution which is sparse and distributed.

One of the simplest approaches to model the findings of the physiology and predict the neural response was the linear systems approach. In the linear systems approach, the sum of the responses to each input is equal to the response to the sum of the inputs. The receptive field, or linear filter, was used as a template to determine the response of a neuron to any novel stimulus. The response was simply computed as the inner product between the receptive field and the stimulus image. This is a linear operation because the response is the weighted sum of the input image pixel intensities. Based on this simple concept, models of the retina, LGN, and V1 simple cells were developed as a single layer of linear filters (Enroth-Cugell and Robson, 1966; Movshon et al., 1978b).

Many of the early studies of V1 neurons tested the linearity of neuron's response. Although a linear model provided a good description of the response to many stimuli, significant deviation from the linearities were also known Movshon et al. (1978a,b); Andrews and Pollen (1979); Tadmor and Tolhurst (1989); DeAngelis et al. (1993); Gardner et al. (1999); Albrecht and Geisler (1991); Reid et al. (1991). One approach to account for these discrepancies was to apply a nonlinear operation to the linear output, such as thresholding function (Tolhurst and Heeger, 1997) or a sigmoidal function of stimulus contrast (Schumer and Movshon, 1984; Tolhurst and Dean, 1987, 1991; Tadmor and Tolhurst, 1989; Albrecht and Geisler, 1991; DeAngelis et al., 1993). However, these linear-nonlinear models still failed to explain the responses observed in V1 to

complex stimuli. For example, response saturation at high contrasts (Albrecht and Hamilton, 1982), and "nonspecific suppression" (Bonds, 1989; DeAngelis et al., 1992; Tolhurst and Heeger, 1997) where the response of a cell to its optimal stimulus is suppressed by the presence of some other stimulus that produces no response when presented alone. In Chapter 3, I will review a wide family of nonlinearities that cannot be explained by the simple linear-nonlinear models.

1.6 The nonlinearities in the visual system and a geometrical framework

Generally, the nonlinearities observed in V1 are analyzed and modeled as separate and independent from each other. These nonlinearities were considered to be something like a"bag of tricks" employed by the visual system where each solves a specific visual problem. For example, the end-stopping nonlinearity is there to detect the ends of the edge. Contrast gain control is to control the gain because neurons have a limited range of response. Each observed nonlinearity has its own mathematical equation and a functional goal. As David Marr once said, "For the subject of vision, there is no single equation or view that explains everything".

In chapter 4, I will describe a new perspective to analyze these nonlinearities. Field and Wu (2004) and Zetzsche et al. (1999), proposed a geometrical framework to look at these different nonlinearities. They argued that by observing the geometry of the neurons nonlinear response, one can gain deep insights into the efficient coding mechanism of the visual system. In Golden et al. (2016) we propose, that many of the nonlinearities can be described by simple curvature in

the geometry of neural response. Similar arguments were made by Zetzsche and colleagues (e.g., Zetzsche et al. (1999); Zetzsche and Nuding (2005)), they too emphasized that many of the nonlinearities observed in V1 can be described by simple curvature of a neuron's response surface. Here, we are not proposing any model that fits all the nonlinearities in the visual system. There have been modeling efforts to unify many of the nonlinearities under a single computational model (e.g., Zhu and Rozell (2013); Mély and Serre (2017)). We (Golden et al., 2016; Vilankar and Field, 2017), are proposing a theoretical and geometrical framework to understand the cause of seemingly different nonlinearities better and argue that a neuron does much and beyond the description depicted by its receptive field.

1.7 Outline of Dissertation

The dissertation is split into eight chapters. Chapter 1 provides an overview of some of the approaches to studying the neural information processing in the visual system and briefly describes the work presented in the dissertation. Chapter 2 describes the psychophysical experiment performed to compute the statistics of different categories of edges in the natural scene images. Based on the statistics, a Bayesian classifier is developed to predict the type of edges and implications are drawn for the early stages of the visual system. Chapter 3 describes the nonlinearities observed in V1 neurons and introduces the concept of image state space and the basic geometry of nonlinearities. Chapter 4 describes how the warping curvature in the response geometry of a neuron describes a wide family of nonlinearities observed in V1. Furthermore, the chapter shows four approaches to produce the warping curvature. Chapter 5 explores

the curvature produced in the response geometry of sparse coding networks and demonstrates that two neurons with the same receptive fields (feedforward weights), can have different curvature in their iso-response contours. Chapter 6 shows that curvature in the response geometry makes a neuron hyperselective and such neurons can break the Gabor-Heisenberg limit. Chapter 7 demonstrates that different learning rules of sparse coding network produce different response geometry which causes neurons to interact with each other in interesting ways. Finally, in Chapter 8 I will conclude the dissertation with some of the possible future directions.

CHAPTER 2

LOCAL EDGE STATISTICS

Many insights into visual processing have been gained by studying the relationship between the neural responses and the statistical regularities of natural scenes (e.g., Field (1987); Atick (1992); Olshausen and Field (1996); Webster and Mollon (1997); Webster and Miyahara (1997); Lewicki and Olshausen (1999); Lewicki and Sejnowski (2000); Geisler et al. (2001); Murray (2013)). The cells in the early stages of the visual system are argued to produce a sparse representation of natural scenes (Field, 1987; Olshausen and Field, 1996). Alternatively, these cells have been interpreted as edge detectors. These edge detectors describe natural scenes as a collection of luminance discontinuities at various orientations and scales. However, the simple cells in V1 do not differentiate between different causes of the edges in natural scenes. A variety of algorithms has been developed to detect edges in images. However, little effort has been put into understanding the statistics of different classes of edges (Balboa and Grzywacz, 2000; DiMattina et al., 2012; Elder et al., 1999; Fowlkes et al., 2007; Ing et al., 2010). Edge detection algorithms (e.g., Canny (1986)) will identify only luminance discontinuities, but it will not identify the cause of an edge (such as an occlusion edge between two objects, a shadow, or a reflectance change edge).

In this chapter, I will investigate the local statistics of edges in relation to the different classes of underlying causes. Figure 2.1 summarizes the different classes of edges. These different classes show that a luminance discontinuity could result from a number of possible causes. First, one object or surface may occlude another. If the illuminations or reflectances of the two surfaces differ, there will be a luminance discontinuity, forming an occlusion edge. Of course,

not all occlusions will create luminance discontinuities. The contrast between the two surfaces may be too low, or if there is a significant texture on the surface, the texture can mask any discontinuity across the surfaces. A second class of edge is the non-occlusion edge, which may result from several causes. Non-occlusion edges can arise from a reflectance change within a surface or from a shadow or illumination boundary; alternatively, a non-occlusion edge may be caused by a change in the surface orientation with respect to the illuminant (e.g., a crease or fold). The discontinuity in luminance may also result from a combination of these effects. In our experiment, we have designed a triangular slider (see Methods section for details) which allows participants to make a soft categorization of edges which have multiple causes.

The studies which have analyzed edge classes have focused only on the occlusion edges (e.g., Balboa and Grzywacz (2000); DiMattina et al. (2012); Hoiem et al. (2011). Some of these studies used the Berkeley Segmentation Database (Martin et al., 2001). These studies relied on human observers to classify edges as occlusion or non-occlusion. Other studies have used an objective measure such as laser range finding to analyze the relationships between depth discontinuities and luminance discontinuities (Howe and Purves, 2002; Huang et al., 2000; Liu et al., 2011; Potetz and Lee, 2003; Yang and Purves, 2003a,b). Whether it is objective techniques (eg., LIDAR) or subjective classification of edges, each technique has its limitations and biases. An objective method such as Light Detection and Ranging (LIDAR) may miss occlusion edges where the depth discontinuities are small in magnitude, but for a human observer that can be a clear occlusion edge (for example a leaf on top of another leaf). It is also difficult to identify other types of non-occlusion edges using a LIDAR.

Similarly, human observers could also miss detecting occlusion edges. For example, human observers might miss an occlusion edge where the local information in the edge is not clear enough to make classification (such as not enough of a luminance discontinuity or similar textures on occluding and occluded objects). To infer the cause of an edge observers require a fair amount of understanding of the 3-D structure of the image. Sometimes, local information is not enough for classification; these edges require more global information at large scales around the edge. McDermott (2004), demonstrated this using a perceptual task to identify junctions (where two edges meet) in images. He first asked participants to identify junctions in full-scale images to obtain ground truth results. Then he compared the participants' performance to identify junctions in local image patches of various sizes. He found that participants were at chance for a 13-pixel diameter (0.25 degree visual angle) patch around the junctions and had more than 90% classification accuracy with a 201-pixel diameter patch.

Similarly, DiMattina et al. (2012) investigated a variety of algorithms and artificial neural networks that classify image patches that contained hand-labeled occlusion edges and image patches that were labeled as within-surface boundaries. The within-surface patches may or may not contain any edge. They found that the accuracy of the algorithms improved with the increasing image patch size. Also, none of the algorithms compared well with the human observer performance, except a neural network combining information across location and scale.

There are algorithms which make use of local edge statistics to identify edges (Canny, 1986; Leclerc and Zucker, 1987; Shashua and Ullman, 1990; Elder, 1999;

Martin et al., 2004; Zhou and Mel, 2008). One of the important steps in a number of computer vision algorithms is to segregate figure from the background (Heitger et al., 1994; Vecera et al., 2002; Fowlkes et al., 2007; Hoiem et al., 2011). Many techniques have been developed to take advantage of the smooth structure of natural scenes to integrate long-range edges and contours (see, e.g., Elder (1999); Geisler et al. (2001); Li and Gilbert (2002)). However, these algorithms do not take into consideration the local statistics of different edge classes which may help locate significant edges and combine these edges into veridical contours and figures.

Here, we investigate the local statistics of different classes of edges in natural scenes.In this study, we used an edge detection algorithm to identify edges in natural images and then had human observers to classify them into different classes of edges. In our study, we focused on the local statistical differences between these edge classes.

2.1 Methods

The text in this section is directly taken from the published article on this study (Vilankar et al., 2014).

This section describes the experimental procedures used to obtain the categorization of edges as occlusions or non-occlusions. Participants also classified non-occlusion edges further into three sub-categories. The categories of edges were:

1. Occlusion: formed when an object partially occludes another object.
2. Non-occlusion:

(a) Reflectance change: formed when there is a change in reflectance due to surface properties.

(b) Surface change: formed when there is a physical angle change on an object's surface.

(c) Cast shadow: formed when an object casts its shadow on another object.

Figure 2.1 shows examples of the occlusion and three sub-categories of non-occlusion edges.

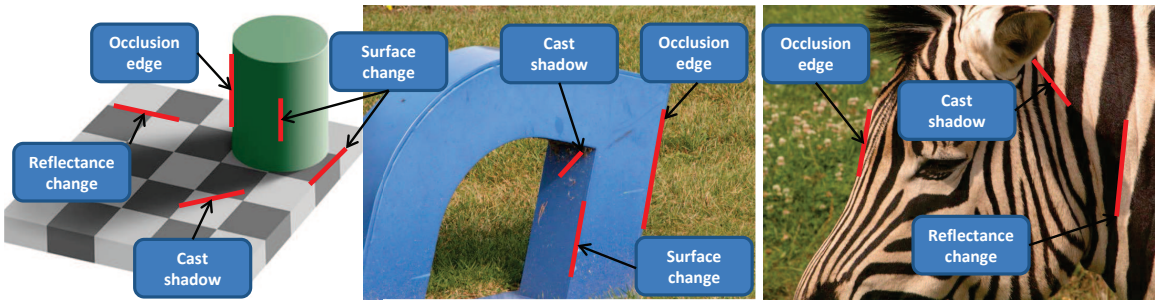


Figure 2.1: Examples of occlusion and non-occlusion edges. The first image is a modified version of Adelson's checkerboard illusion. (http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html).

2.1.1 Apparatus

Stimuli were displayed on an HP LP2465 widescreen LCD monitor. The screen size of the monitor was 52.0×32.6 cm (width × height) with a display resolution of 41 pixels/cm and a frame rate of 60 Hz. The display had a minimum, maximum and mean luminance of 0.38, 350, and 76.5 cd/m^2, respectively, and an overall gamma of 2.2. Stimuli were viewed binocularly through natural pupils in a darkened room at a distance of approximately 60 cm.

2.1.2 Stimuli

Stimuli were generated from thirty-eight high-resolution (2560×1920) natural images from the McGill Color Image Database (Olmos and Kingdom, 2004). The images were selected from seven of the nine categories of the McGill Color Image Database: Flowers, Animals, Foliage, Fruits, Landscapes, Winter and Shadows. No images were selected from the Textures and Man-made categories. The selected images were typically dominated by a small number of objects (as shown in Figure 2.2) which made the process of hand tracing the edges of objects more straightforward (see section 'Human labeled occlusion edges'). However, we also recognize that this selection may produce some biases in our data (see Discussion). The images were displayed in grayscale with 8-bit resolution and pixel values from $0 - 255$. For each image, edges were located using Matlab's Canny edge detection algorithm. The standard deviation of the Gaussian filter used by the Canny algorithm was set to 10 pixels. The low and high thresholds were automatically selected by the Canny algorithm for each image.

A set of 1000 edges found by the Canny algorithm were selected randomly from the 38 natural images. The selected edge locations were uniformly distributed over the image area. No two selected edge locations in an image were within a distance of 80 pixels of each other.

Using the selected edges, 1000 image stimuli were generated. To generate each stimulus, a red bounding box was placed around a selected edge in an image. The 100×100-pixel bounding box subtended a visual angle of approximately 2.4 degrees. The entire stimulus subtended a visual angle of 36.6 degrees. Along with the red bounding box, the edge line from the Canny algorithm was also placed on top of the actual edge in red. Figure 2.3 shows the graphical user

Figure 2.2: Images from the McGill color image database used in the study.

interface with the stimuli.

2.1.3 Participants

Six graduate student volunteers (mean age = 27 years) from the Computational Perception and Image Quality Laboratory at Oklahoma State University took part in the experiment. The experiment was approved by the Institutional Review Board of Oklahoma State University. The participants were given two weeks to complete the experiment in five sessions. The sixth participant completed four of the sessions in a single day, and his results contained many outliers when compared with the other participants. As a result, his responses were excluded from the analysis.

2.1.4 Procedure

Each subject categorized the edge type of all 1000 stimuli with a slider in the user interface. At the onset of a stimulus, a red bounding box with a cross-hair and a red line over the target edge flashed alternately on and off in one-second intervals until a response was made by the subject. Participants observed the selected edge with and without the red bounding box and responded as to whether the displayed edge was an occlusion edge or a non-occlusion edge using a slider. The extreme left of the slider represented 100% confidence that the displayed edge was an occlusion edge, whereas the extreme right represented 100% confidence that the displayed edge was a non-occlusion edge. Figure 2.3 shows the user interface for the experiment. The top figure shows the interface with the horizontal slider.

Edges rated to be non-occlusion edges with a confidence of 75% or higher were further divided into three sub-categories. In order to make a non-occlusion categorization, participants were shown a new triangular-shaped slider on the screen after they made the occlusion vs. non-occlusion response. The bottom image of Figure 2.3 shows the user interface with the triangular selector and a red cross-hair within that triangle. The three vertices of the triangle represented unambiguous judgments (100%) of the non-occlusion sub-categories. Participants made their judgment of the non-occlusion subcategory by placing the red cross-hair at an appropriate position in the triangle. For example, when the red cross-hair was placed near the vertex representing a reflectance change edge, the user was indicating a very high confidence that the highlighted edge was a reflectance change. If the red cross-hair was placed midway between any two vertices of the triangle, the subject judged the edge to have equivalent proper-

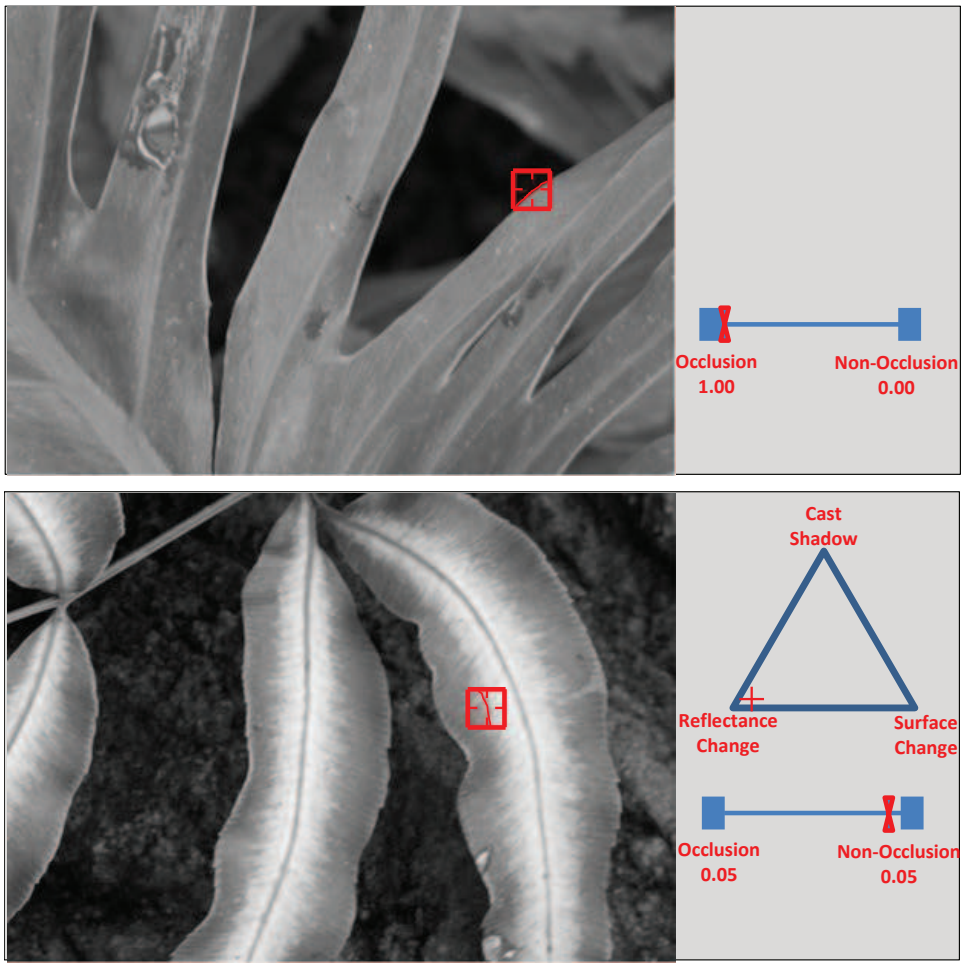


Figure 2.3: The graphical user interface for the experiment. The image on top shows the interface used to make a decision between occlusion and non-occlusion categories using the horizontal slider. The image below shows the interface with the triangular slider used for the sub-categorization of non-occlusion edges after the participant has rated the edge to be in the non-occlusion category with a confidence of 75% or higher.

ties of the two edge categories represented by the two vertices. Similarly, if the red cross-hair was placed at the center of the triangle, the subject judged the edge to have properties from all non-occlusion sub-categories.

2.1.5 Potential biases

The methodology we introduce here has several potential sources of bias. Here, we wish to emphasize three potential biases. However, it is also worth noting that any technique that attempts to deduce the underlying cause of an edge will suffer from some bias as there is no objective measure of ground truth. In our methods, we used human observers and a Canny edge detection algorithm to detect and classify edges. These are imperfect methods. The use of human observers certainly introduces potential biases. The instructions to observers, the choice of images and the choice of parameters in the Canny operator are all likely to have some effect on the statistics described here. Although we believe we have selected a reasonable set of parameters, it will not be clear what effects they have until a large variety of studies are performed that explore the space of parameters. The three choices we wish to emphasize are:

1. The choice of images: We selected images from the McGill database that had well defined objects with reasonably well defined boundaries. A much larger database of images needs to be explored. Images like that of the Van Hateren image set van Hateren and van der Schaaf (1998), for example, contain many scenes where edges are quite difficult to label (e.g., grass, leaves) where many of the edges are from objects that approach the sizes of the pixels. We are currently exploring how the image set affects these statistics.

2. The Canny edge detection algorithm has several settings: For example, we chose a particular scale for most of these studies (a 10-pixel scale Gaussian filter). We have repeated a portion of these studies with a larger scale (a 20-pixel scale Gaussian filter) and see largely similar results. However,

there is large space of parameters that could be explored and we cannot be confident that a different choice parameters will not alter these results. We do believe however that the parameters we chose represent a reasonable first attempt.

3. Procedures and observers: Our method of classifying edges began with the classification of occlusion versus non-occlusion and then proceeding to a three-way classification of non-occlusion edges. Although we found that this approach was reasonable, it is not clear how different procedures might alter these results. It should be noted that Elder et al. (1999), in an unpublished study, produced largely similar classification results with different procedures.

2.2 Results

The text in this section is directly taken from the published article on this study (Vilankar et al., 2014).

2.2.1 Occlusion edges in natural scenes

Across participants, approximately 50% of the edges were classified as occlusion edges and 50% of the edges were classified as non-occlusion edges. If a participant rated an edge as an occlusion with 50% (or higher) confidence, then that edge was classified as an occlusion edge for this measure. Individually, Participants 1 to 5 identified 49%, 50%, 52%, 51%, and 51% of edges as occlusion edges, respectively. Overall, 50.6% of the edge stimuli were classified as occlu-

sion edges. In addition, the results for the experiment repeated with a larger scale (20-pixel) Canny operator yielded very similar proportions of occlusion edges: overall, 49% of the edge stimuli were classified as occlusion edges.

We also examined the degree of mutual agreement between the five participants. Table 2.1 shows the proportion of occlusion edges and non-occlusion edges with between-participant agreements of 100% (5 out of 5 participants), 80% (4 out of 5), and 60% (3 out of 5). The fact that the proportions are similar across different degrees of mutual agreement indicates that, for edges found by the Canny edge detector, occlusion edges occur as frequently as non-occlusion edges. The fourth column in the table shows the proportion of the edges which did not satisfy the minimum mutual agreement criteria. This column indicates that for 100% between-participant agreement, classifications for 13% of the edge stimuli did not meet the criterion (i.e., agree across all 5 participants). For 80% between-participant agreement, only 5% of the edge stimuli did not meet the criterion.

2.2.2 Analysis of occlusion vs. non-occlusion edges

Next, we examined the local statistical properties of edges based on their classification as occlusions or non-occlusions. Images were analyzed using linear luminance values.[1]

To analyze the statistics of the edges, we selected only those occlusion and non-occlusion edges which had at least 80% mutual agreement (4 out 5 participants agreed on the category). 946 edges out of 1000 edges had 80% between-

[1]We also analyzed edges from log-luminance images and found similar results.

Table 2.1: The proportion of occlusion and non-occlusion edges at various degrees of mutual agreement between-participants. The fourth column shows the proportion of edges which did not satisfy the between-participant agreement criteria.

Mutual agreement	**Occlusion edge**	**Non-occlusion edge**	**Edges not satisfying the criterion**
100% (5 out of 5)	44%	43%	13%
80% (4 out of 5)	48%	47%	5%
60% (3 out of 5)	50%	50%	0%

participant mutual agreement. However, we selected only 673 edges (330 occlusion edges and 343 non-occlusion edges) for the edge patch extraction. The remaining 273 edge patches (145 occlusion edges and 128 non-occlusion edges) were not selected because those patches had more than one edge within the patch. The selected edges were then extracted into small patches of 81×41 pixels. These patches were aligned using the Radon transform such that the edge line was oriented horizontally and located at the center of the patch at the 41^{st} pixel row. The patches were additionally oriented such that the higher-luminance half of the patch was always placed on top and the lower-luminance half of the patch was placed on bottom. Figure 2.4 (a) shows an illustration of an extracted patch, which has higher- and lower-luminance areas separated

by an edge. All the extracted edge patches are freely available from our online database (http://redwood.psych.cornell.edu/edges).

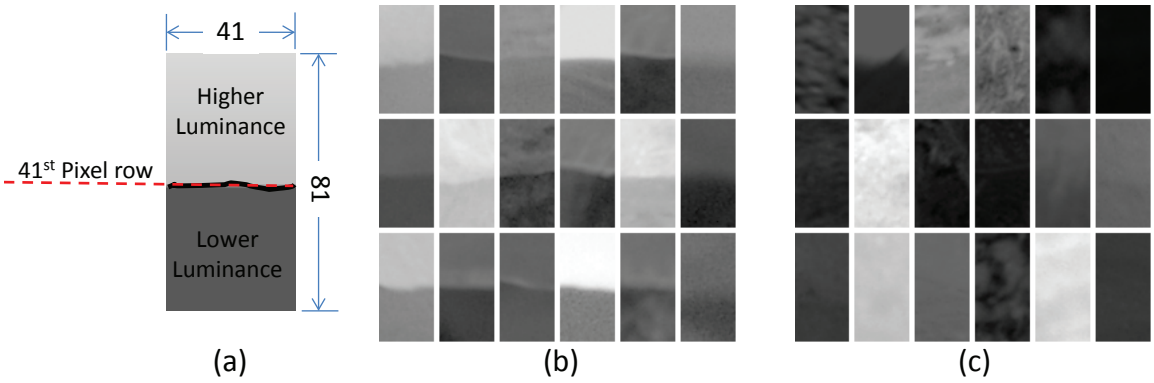


Figure 2.4: a) An 81×41-pixel extracted edge patch . The patch is aligned such that the higher-luminance side is on top and the lower-luminance side is on the bottom. The edge line is between the two sides is at the 41^{st} pixel row. (b) A sample of the extracted occlusion edges. (c) A sample of the extracted non-occlusion edges. Both sets of extracted edges were first identified using the Canny edge operator and then classified by human observers.

Figure 2.4 (b) shows a set of extracted occlusion edges and Figure 2.4 (c) shows a set of extracted non-occlusion edges. Figure 2.5 show some of the patches not selected for the statistical analysis of occlusion and non-occlusion edges.

The contrast distribution of occlusion vs. non-occlusion edges

We measured the distribution of contrasts for edges classified as occlusions and non-occlusions with both Michelson contrast and root mean square (RMS) contrast. Michelson contrast measures the contrast between the two sides of the occlusion edges (occluding side and occluded side), whereas RMS contrast mea-

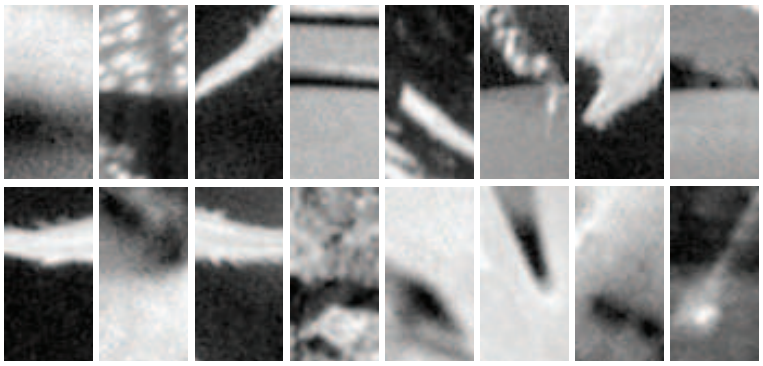


Figure 2.5: A set of edge patches not selected for the statistical analysis of occlusion and non-occlusion edges. These patches have multiple edges in the extracted patch.

sures the contrast over the entire edge patch. The Michelson contrast for an edge patch was calculated as follows:

$$C_m = \frac{L_{top} - L_{bottom}}{L_{top} + L_{bottom}} \tag{2.1}$$

where L_{top} is the mean luminance of the higher luminance (top) section, and L_{bottom} is the mean luminance of the lower luminance (bottom) section of the edge patch. To compute the L_{top} and the L_{bottom}, the 81×41 edge patch was divided into three sections. The sizes of the top , middle, and bottom sections were 30×41, 21×41, and 30×41 respectively. The L_{top} and the L_{bottom} were computed using the top and the bottom sections, respectively.

$$\begin{aligned} L_{top} &= \tfrac{1}{41\times30}\textstyle\sum_{x=1}^{41}\sum_{y=1}^{30} ep(x,y) \\ L_{bottom} &= \tfrac{1}{41\times30}\textstyle\sum_{x=1}^{41}\sum_{y=52}^{81} ep(x,y) \end{aligned} \tag{2.2}$$

where $ep(x, y)$ denotes the luminance value at x^{th} and y^{th} pixel location in the edge patch. The middle section which included the edge line was excluded from the Michelson contrast computation to reduce the effects of edge blur and

edge curvature.

RMS contrast was measured as the ratio of the standard deviation to the mean luminance of the edge patch.

$$C_{RMS} = \frac{\frac{1}{41\times 81}\sqrt{\sum_{x=1}^{41}\sum_{y=1}^{81}(ep(x,y)-\overline{ep})^2}}{\overline{ep}} \tag{2.3}$$

where $\overline{ep}$ denotes the mean luminance of the patch and x and y denote pixel coordinates.

Figure 2.6 (a) - (d) show the histograms for the contrast of occlusion and non-occlusion edge patches. Figure 2.6 (a) and (b) show the Michelson contrast histograms, and (c) and (d) show the RMS contrast histograms. The horizontal axis of each histogram specifies the contrast of an edge patch, and the vertical axis specifies the number of edges in that contrast range. These distributions reveal that for those edges found by the Canny algorithm, the edges classified as occlusion by participants have a relatively high contrast compared to the edges classified as non-occlusion edges. Similar distinctions were observed between the log-luminance occlusion and non-occlusion edge patches (not shown).

Figure 2.6 (e) and (f) show the empirical cumulative distributive functions (CDFs) for Michelson and RMS contrasts of occlusion and non-occlusion edges. Figure 2.6 (e) shows the CDF for the Michelson contrast of edges, and (f) shows the CDF for the RMS contrast. The horizontal axis represents the contrast of edges and the vertical axis represents the proportion of edges with a contrast below that indicated on the horizontal axis. As can be seen from the CDF of the Michelson contrast, 75% of the occlusion edges have contrast values more than 0.42, and 75% of the non-occlusion edges have contrast values less than

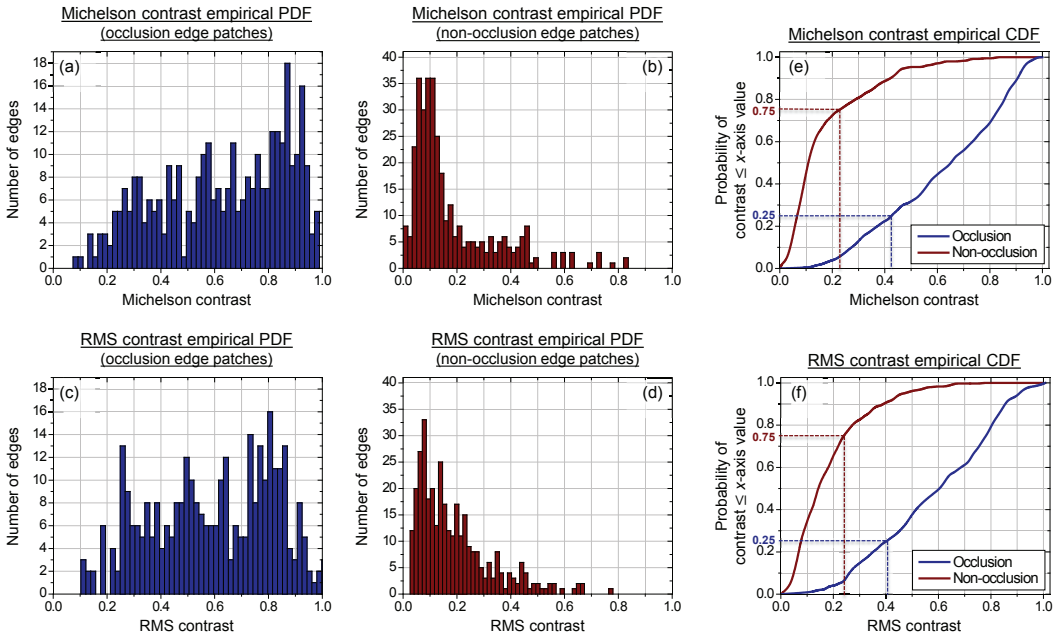


Figure 2.6: The histograms of Michelson contrast and RMS contrast for occlusion and non-occlusion edges. (a) and (b) show the histogram of Michelson contrast in occlusion and non-occlusion edge patches. Similarly, (c) and (d) show the histograms of RMS contrast in occlusion and non-occlusion edge patches. (e) and (f) show the empirical CDF of Michelson contrast and RMS contrast in occlusion and non-occlusion edge patches. The blue curve shows the CDF of contrast in occlusion edges and the red curve shows the CDF of contrast in non-occlusion edges.

0.23. Similarly from the CDF of the RMS contrast, 75% of occlusion edges have contrast values more than 0.41, and 75% of non-occlusion edges have contrast values less than 0.25. Similarly, for the log-luminance edge patches (not shown), 75% of the occlusion edges had Michelson contrast values more than 0.075, and 75% of the non-occlusion edges had contrast values less than 0.035; for RMS contrast, 75% of occlusion edges had contrast values more than 0.071, and 75% of non-occlusion edges had contrast values less than 0.037. These results indicate that Michelson contrast or RMS contrast can be used as a strong cue in predicting whether an edge located by the Canny algorithm will be classified as

occlusion or non-occlusion edge by participants.

The average occlusion and non-occlusion edges

We computed the average normalized occlusion edge and non-occlusion edge. First, each edge patch was normalized such that it spanned the range from $0-1$. The average occlusion and non-occlusion edge patches were then computed as follows:

$$\mu(x,y) = \frac{\sum_{i=1}^{N} ep_i(x,y)}{N} \tag{2.4}$$

where, $\mu(x,y)$ is the average luminance at pixel location x and y, ep_i denotes the i^{th} extracted patch, N denotes the total number of the extracted edge patches, and x and y denote the pixel coordinates.

The two-dimensional average edge patch was then converted to a one-dimensional average edge profile by averaging across each row in the two-dimensional patch. Figure 2.7 (a) and (b) show the one-dimensional average occlusion and non-occlusion edges, respectively. These data show that the average occlusion edge has a sharper transition from low luminance to high luminance than the average non-occlusion edge. Also shown are a sample of 20 randomly selected occlusion or non-occlusion edges plotted in blue. The sample occlusion edges clearly have a steeper transition than the sample non-occlusion edges. The average normalized log-luminance occlusion and non-occlusion edges (not shown) yielded similar results.

We also compared the slopes of luminance transition for occlusion and non-

occlusion edges. For each extracted two-dimensional edge patch, a slope was computed by first converting the edge patch to a one-dimensional edge profile with a length of 81 pixels. Then the slope of each one-dimensional edge profile was computed as a mean change in the luminance from the 36^{th} pixel to the 46^{th} pixel. The slope of transition of occlusion edges was significantly higher ($t(671) = 16.08$, $p < 0.0001$) than the slope of non-occlusion edges. The average non-occlusion edge appears to be non-monotonic with the greatest contrast difference near the center of the edge. We will return to this point later in describing the average edges of non-occlusion sub-categories.

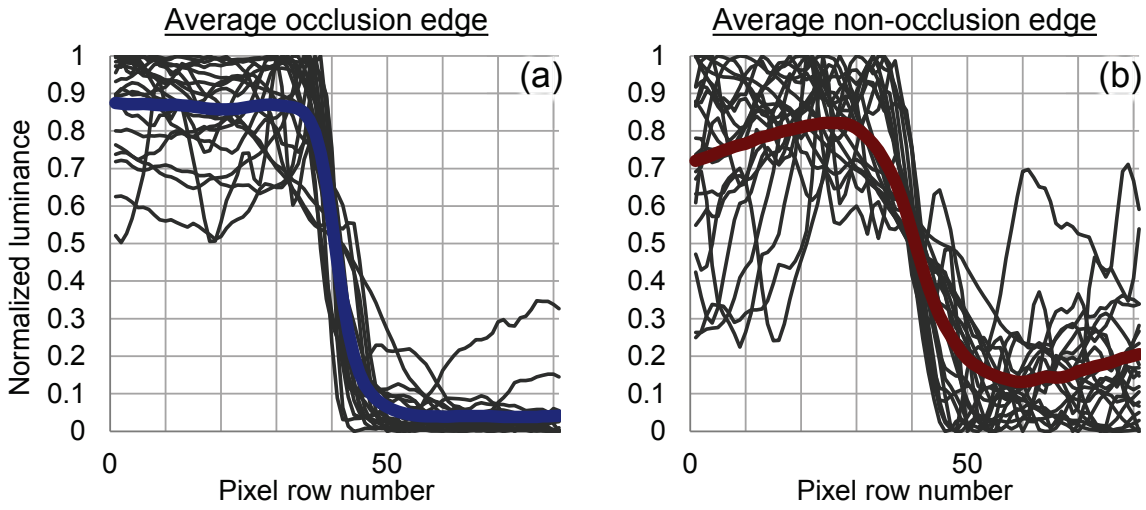


Figure 2.7: One dimensional profiles of normalized average occlusion and non-occlusion edges. (a)The normalized average occlusion edge in blue with 20 sample occlusion edges. (b) The normalized average non-occlusion edge in red with 20 sample non-occlusion edges. The edges in (a) and (b) were first detected by the Canny operator and then categorized by participants as occlusion or non-occlusion edges. The slope of occlusion edges was significantly different ($t(671) = 16.08$, $p < 0.0001$) from the slope of non-occlusion edges.

Mean luminance vs. contrast

Here, we investigate the relationship between contrast and mean luminance of the edge patches. Mante et al. (2005) found that contrast and luminance were statistically independent. The image patches they analyzed were extracted from a simulated saccadic inspection of natural scenes. Their patches did not necessarily include an edge. To determine whether our edge patches also showed this independence, we analyzed the relationship between contrast and mean luminance for our extracted edge patches. Michelson contrast was computed as shown in Equation 2.1 and RMS contrast was computed as shown in Equation 2.3. Mean luminance was computed as follows:

$$L_{mean} = \frac{L_{top} + L_{bottom}}{2} \tag{2.5}$$

where L_{top} is the mean luminance of the higher-luminance top section of the patch and L_{bottom} is the mean luminance of the lower-luminance bottom section of the patch.

Figure 2.8 (a) and (b) show the scatter plots of mean luminance vs. Michelson contrast and mean luminance vs. RMS contrast of edge patches, respectively. The blue circles represent points from the occlusion edge patches, and the red circles represent points from the non-occlusion edge patches. As can be seen from Figure 2.8 (a) and (b), there is no significant correlation between the contrast and the mean luminance, except the correlation ($r(341) = -0.13$, $p = 0.02$) between the Michelson contrast vs. mean luminance for non-occlusion edges which has a weak but significant correlation. These results are consistent with the findings of Mante et al. (2005) and suggest that mean luminance and contrast are largely independent for both occlusion and non-occlusion edges. One

should note that the measure of RMS contrast used here is a normalized RMS contrast, where the RMS contrast was normalized by dividing by the mean luminance of the edge patch (see Equation 2.3). One would expect to have a strong correlation between the non-normalized RMS contrast and the mean luminance as there would be more variation in luminance in the higher mean-luminance edge patches compared to the lower mean-luminance edge patches. Indeed, we found strong correlations ($r > 0.85$) between the non-normalized RMS contrast and the mean luminance for occlusion and non-occlusion edges.

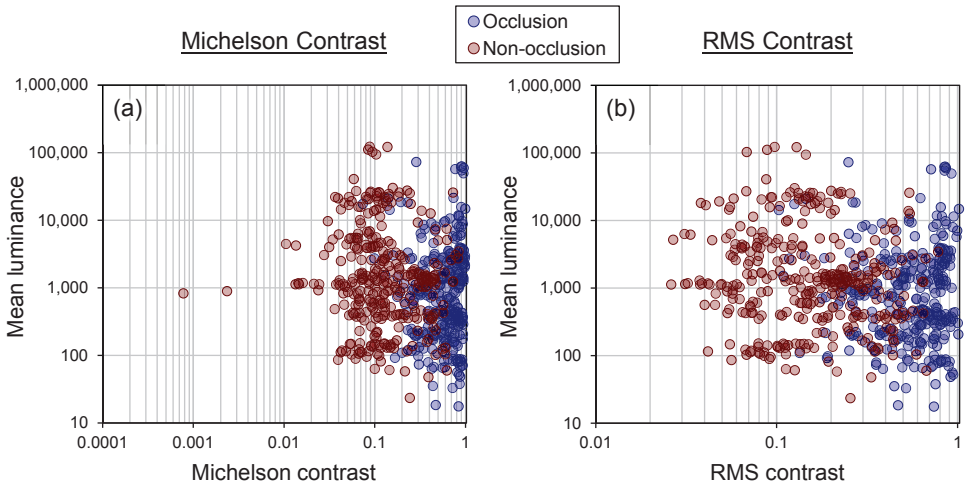


Figure 2.8: Scatter plots of the mean luminance vs. contrast of occlusion and non-occlusion edge patches. (a) shows the scatter plot of mean luminance vs. Michelson contrast of occlusion (correlation $r(328) = 0.08$, $p = 0.15$) and non-occlusion (correlation $r(341) = -0.13$, $p = 0.02$) edge patches. (b) shows the scatter plot of mean luminance vs. RMS contrast of occlusion (correlation $r(328) = 0.08$, $p = 0.15$) and non-occlusion (correlation $r(341) = -0.10$, $p = 0.06$) edge patches. The blue open circles represent occlusion edge patches and the red open circles represent non-occlusion edge patches.

2.2.3 Analysis of non-occlusion edge sub-categories

All of the edges labeled as non-occlusion edges with a slider rating of more than 75% were sorted into three sub-categories: reflectance change (RC), cast shadow (CS), or surface change (SC). As mentioned previously, the sub-category rating for each non-occlusion edge was made using a triangular slider where the vertices represented the three sub-categories of non-occlusion edges. The sub-category rating was registered by placing the cross-hair in the triangle at the appropriate position. Figure 2.9 shows the density maps of cross-hair placement in the triangular slider for each participant. Qualitatively, the density map from each participant indicates that most of the non-occlusion edges in the natural scenes used in this study are judged as due to reflectance changes and surface changes. Most of the slider positions lie on the line segment between the vertices corresponding to RC and SC.

Proportion of non-occlusion sub-categories

Figure 2.10 shows the relative proportions of the three sub-categories of non-occlusion edges for each participant. The triangular slider was divided into three regions as shown in the figure. The three sub-regions of the triangle represent the sub-categories of non-occlusion edge. The upper region represents a cast shadow (CS) edge. If the slider cross-hair was placed in this region, then that edge was considered as a cast shadow edge. Similarly, the lower left is the region representing a reflectance change (RC) and the lower right region represents a surface change (SC) edge. The relative proportions of sub-categories of non-occlusion edges per participant were significantly different [$F(2, 12) = 12.11$, $p = 0.0013$]. The Tukey HSD post hoc test indicated the cast

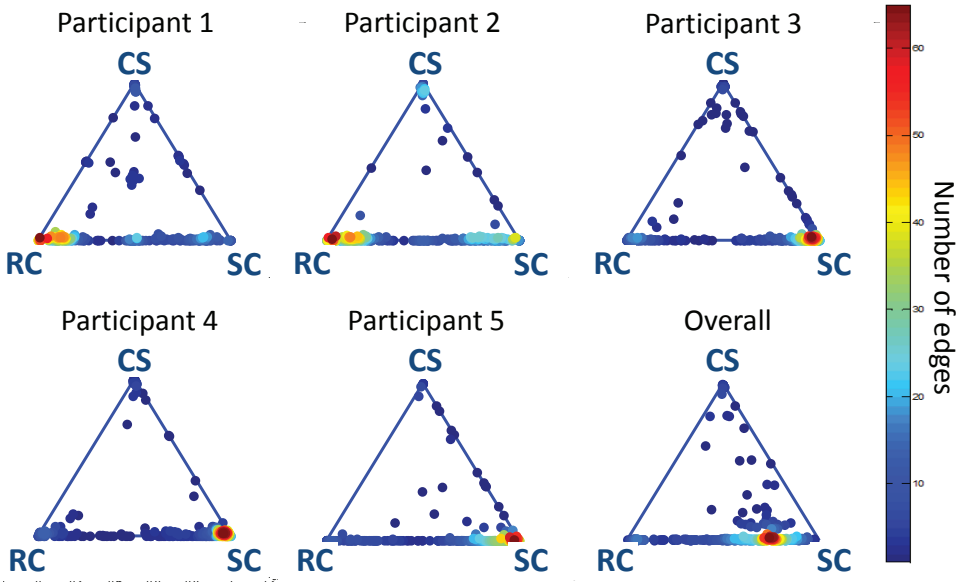


Figure 2.9: The density of the triangular slider placement for each participant and the overall average density of placement of the slider from the mean positions of the slider across participants for each edge. Note: The overall density is not the sum of slider position of all participants. Each point corresponding to an edge in the overall density map is the result of averaging the position slider for that edge across all participants (Vertices: Reflectance Change (RC), Cast Shadow (CS), and Surface Change (SC)).

shadow edges ($M = 40$, $SD = 8$) occur significantly less frequently than surface change edges ($M = 269$, $SD = 102$). However, the relative proportions of reflectance change edges ($M = 151$, $SD = 76$) were not significantly different from cast shadow or surface change edges. Here M and SD represent mean relative proportion and standard deviation. Additionally, most of the non-occlusion edges were classified as reflectance changes and surface changes. However, the ratings were not consistent across participants. Participants 1 and 2 categorized non-occlusion edges predominantly due to reflectance changes, while participants 3, 4 and 5 categorized non-occlusion edges as mostly due to surface changes. Overall, approximately 31% of edges were due to reflectance changes,

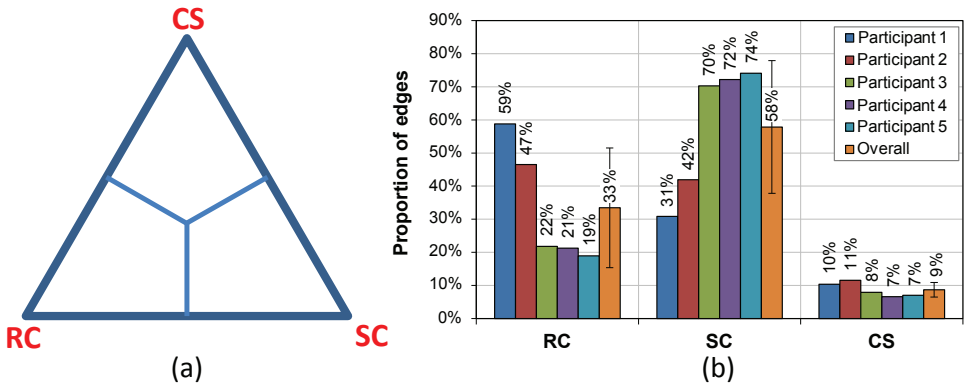


Figure 2.10: (a)The triangular slider for sub-categorization was divided into three regions representing the three sub-categories (Reflectance Change (RC), Cast Shadow (CS), and Surface Change (SC)) of non-occlusion edges. (b)The proportions of sub-categories of non-occlusion edges for each participant and overall mean proportions. Overall, approximately 31% of edges were due to reflectance changes, 8% were due to cast shadows, and 56% were due to surface changes.

8% were due to cast shadows, and 56% were due to surface changes. In addition, the results for the experiment repeated with the larger scale (20-pixel) Canny operator yielded very similar proportions of non-occlusion edges: overall, approximately 25% of edges were to due to reflectance changes, 9% were due to cast shadows, and 66% were due to surface changes.

We also examined the degree of mutual agreement between the five participants. Figure 2.11 (a) shows the absolute proportions of the three categories of non-occlusion edges at 60%, 80%, and 100% between-participant mutual agreement. That is, it shows agreement between 5 out of 5 participants, 4 out of 5 participants, and 3 out of 5 participants, respectively. Figure 2.12 shows a sample of edges in each sub-category with a red bounding box around the edges. All of the edges shown in the top three rows corresponding to the three sub-categories have at least 80% between-participant agreement; the edges in the

bottom row are the indeterminate edges which did not meet the 80% between-participant agreement. Overall, the cast shadow edges occur rarely as compared to the other two categories, while surface-change edges occur most frequently. However, there is a large variation in the proportion of surface change edges at different between-participant mutual agreements. With 60% mutual agreement, the proportion of surface change edges is 70%; with 100% mutual agreement the proportion of surface change edges is only 8%. This suggests that high disagreement on edges classified as surface changes led to many edge patches that did not meet the 80% mutual agreement criterion. Similar variations can be seen for the reflectance change edge proportions. Figure 2.11 (b) shows these variations in the sub-regions of the triangle. The three ellipses in the figure represent the variations between participants. The major axis and the minor axis of the ellipses represent the standard deviation in the horizontal and the vertical directions, respectively. The asterisk at the center of each ellipse represents the overall mean of the triangular slider placements of the five participants in each sub-region, and the circular dots coded with different colors show the density of the mean slider placement.

Local statistics of non-occlusion edges

The average Michelson contrasts of non-occlusion edges in each category at different degrees of mutual agreement are shown in Figure 2.13. We found significant differences between the contrast of edges in each category [$F(2, 193) = 99.92$, $p < 0.0001$]. A post-hoc comparison using the Tukey HSD test also revealed a significant difference in contrast between each category. The cast shadow edges exhibit the highest Michelson contrast ($M = 0.616$, $SD = 0.19$)

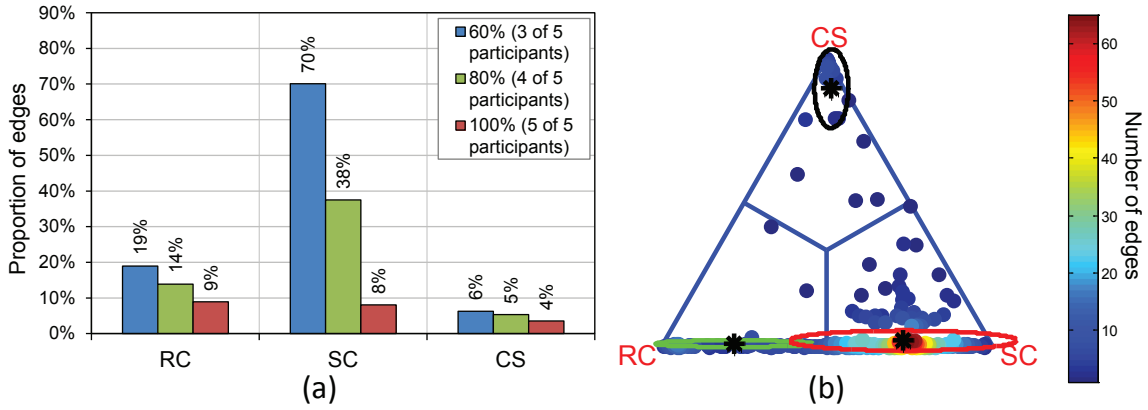
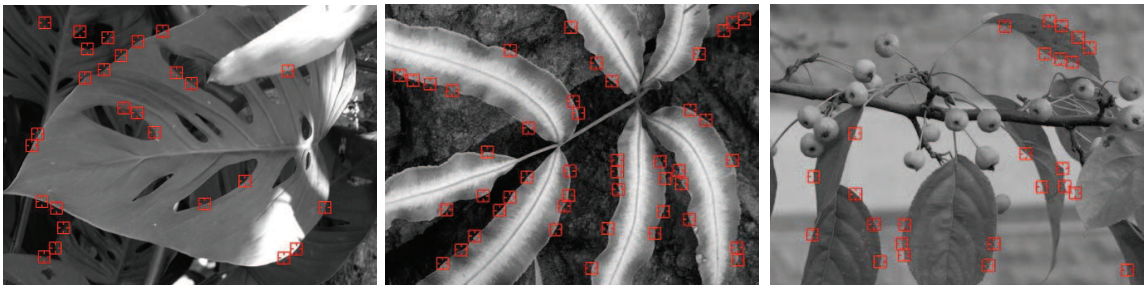


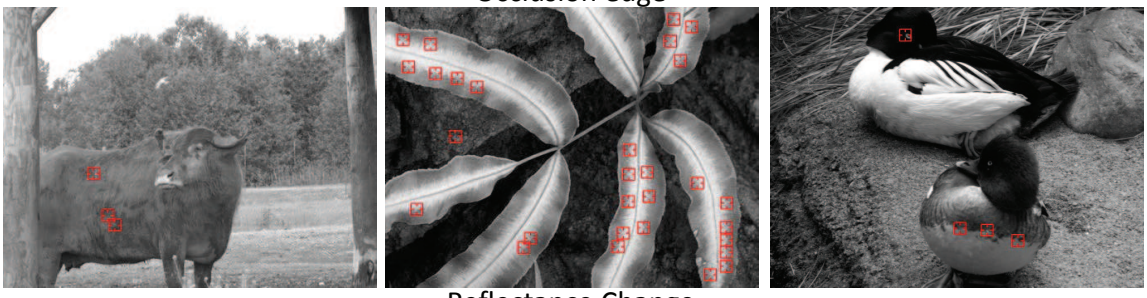
Figure 2.11: (a) The proportions of sub categories of non-occlusion edges at different degrees of mutual agreement between participants. (b) The three ellipses show the variations in horizontal and vertical directions in each sub-region corresponding to the three sub-categories of non-occlusion edges. The three asterisks represent the mean placement of the slider for all edges in each sub-region. The colored circles represent the mean placement of the slider for each edge.

and the surface change edges exhibit the lowest Michelson contrast ($M = 0.142$, $SD = 0.113$) amongst all non-occlusion edge sub-categories. Here M and SD represent the mean and standard deviation of contrast.

Figure 2.14 (a), (b) and (c) show the normalized average edge for each sub-category of non-occlusion edges. These data indicate that the normalized average cast shadow edge and surface change edge have sharper transitions from lower luminance to higher luminance. We found significant differences in the slopes of non-occlusion subcategories [$F(2, 193) = 8.15$, $p = 0.0004$]. Post-hoc comparisons using the Tukey HSD test indicated that mean slope of the cast shadow edges ($M = -0.042$, $SD = 0.023$) was significantly different from mean slope of reflectance change edges ($M = -0.022$, $SD = 0.01$) and surface change edges ($M = -0.026$, $SD = 0.015$). However, the mean slope of reflectance change

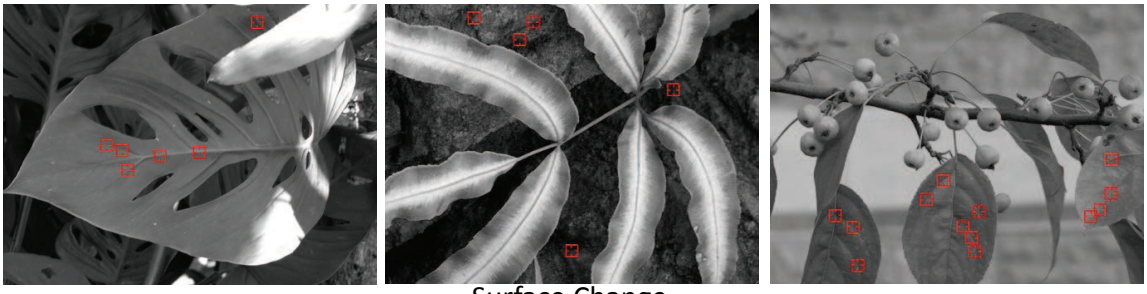


Occlusion edge

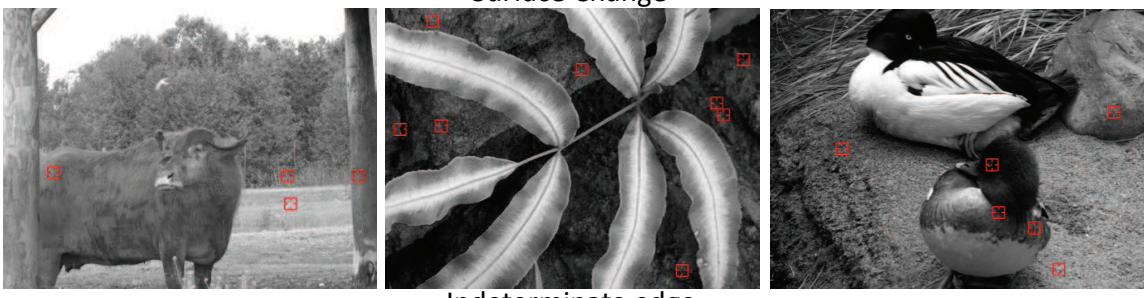


Reflectance Change

Cast Shadow



Surface Change



Indeterminate edge

Figure 2.12: Samples of occlusion and non-occlusion edges categorized with at least 80% between-participant mutual agreement. Each edge shown here is bounded by a red box. The first row shows edges categorized as occlusion edges. The next three rows correspond to edges categorized as reflectance changes, cast shadows, and surface changes, and the bottom row shows the indeterminate edges which did not meet 80% between-participant mutual agreement.

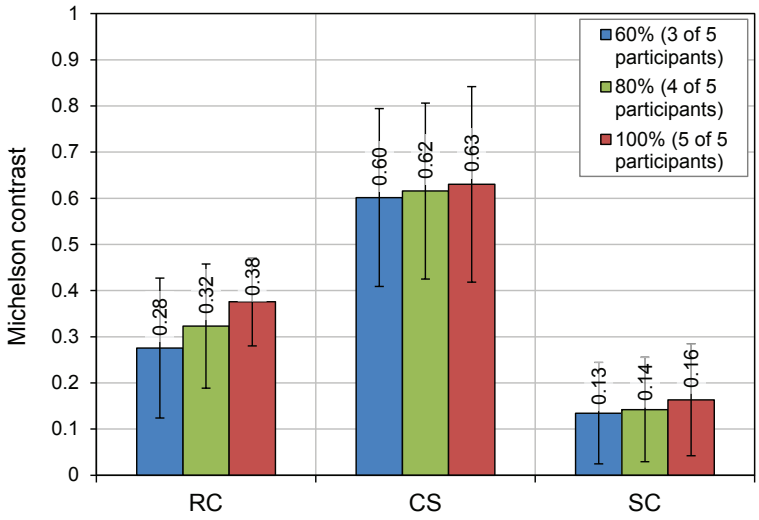


Figure 2.13: The mean contrast of each sub-category at different degrees of between-participant mutual agreement. The error bars represent the mean standard deviations of contrast in each category. The contrast difference between each category was statistically significant [$F(2, 193) = 99.92$, $p < 0.0001$].

and surface change edges were not significantly different. Here M and SD represent the mean and standard deviation of the slope of normalized average edges.

As we noted earlier, the average non-occlusion edge shows a non-monotonic luminance profile. Figure 2.14 (c) shows this result is primarily due to the surface change sub-category. Further analysis is required, but we speculate that this profile results from a peak in the reflection at the center of the fold in the surface. This may be due to a very local change in the shape of the surface. As

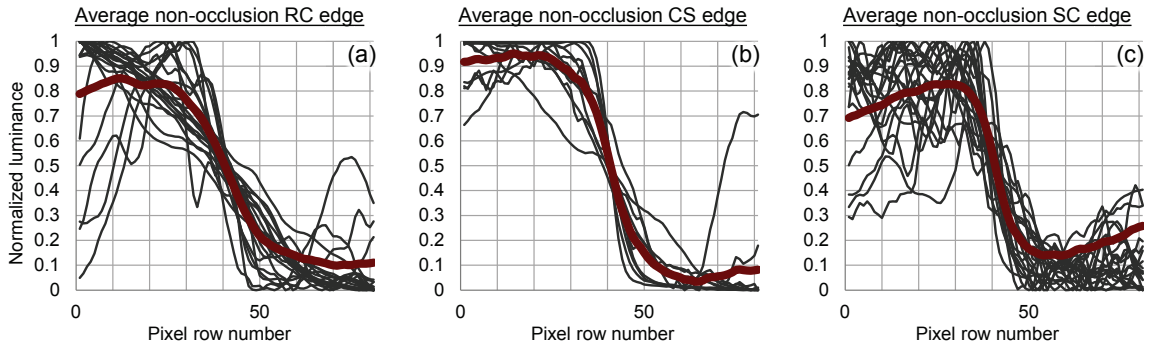


Figure 2.14: One-dimensional profiles of normalized average non-occlusion edge sub-categories. (a) The normalized average reflectance change (RC) non-occlusion edge in red with 20 sample occlusion edges. (b) The normalized average cast shadow (CS) non-occlusion edge in red with 13 sample occlusion edges. (c) The normalized average surface change (SC) non-occlusion edge in red with 20 sample occlusion edges. The slope of CS edges were significantly different from RC ($p < 0.01$) and SC ($p < 0.01$) edges.

distance from the local surface change increases, the reflected intensity returns to the mean intensity of the surface.

2.2.4 Human-labeled occlusion edges

It is important to note that the results presented in previous sections may have an intrinsic bias, as they were based on the edges found by the Canny detection algorithm. The Canny algorithm determines that an edge is present when there is a luminance difference above a certain threshold (Canny, 1986). In order for an edge to be classified as an occlusion edge in the experiment, it must first be detected by the Canny algorithm. Based on the results of DiMattina et al. (2012), it is likely that many occlusion boundaries easily identified by participants will be missed by the Canny operator. This can be true for both texture edges and

low contrast edges that have long range support (that is, they can be inferred by integrating along the contour). To construct a broader account of occlusion, we asked participants to identify the locations of occlusion edges by tracing these edges on our collection of images.

Edge tracings

Three participants were asked to trace the occlusion edges in 38 natural scene images from the McGill Color Image Database. The images were displayed in their original color versions. Their instructions were as follows: *In the displayed image you should only trace the edges which occur when an object occludes another object. Only trace the edges which are formed by the main objects of the images. Ignore the occlusion edges formed by small objects such as: grass, leaves, small flowers, etc.* Participants were shown examples of tracings done earlier by the first author of this paper. Participants used the Adobe Photoshop Brush tool controlled by a mouse for tracing on color versions of the natural images. The Brush tool was set to a diameter of nine pixels and the color red. The edges were traced on a separate Adobe Photoshop layer and overlaid on the top of the image layer. The left side of Figure 2.15 shows an image displayed to a participant for the tracing of the occlusion edges and the right side shows the resulting tracing in red.

Using the occlusion edge traces from the participants, we extracted 81×41-pixel edge patches. These patches were oriented using the same procedure as the edge patches extracted using the Canny algorithm. Figure 2.16 shows samples of the occlusion edge patches extracted using the hand-traced data. All the extracted edge patches are freely available from our online database (http://redwood.psych.cornell.edu/edges).

(a) (b)

Figure 2.15: (a) Original high-resolution image and (b) occlusion edge tracing in red from a participant.

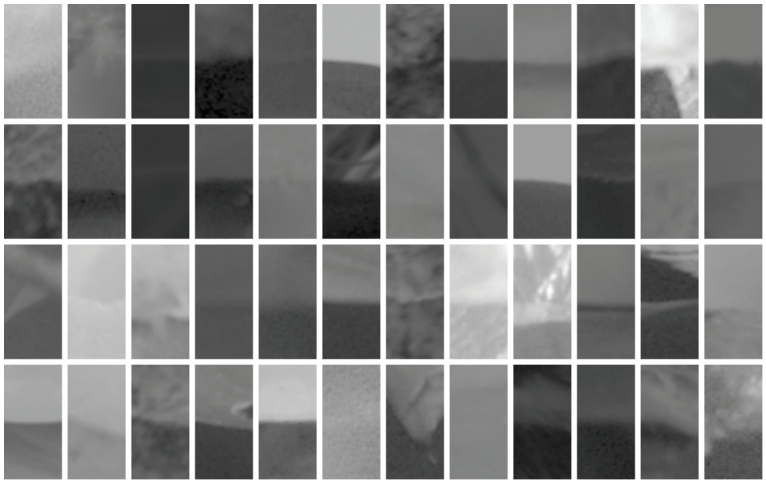


Figure 2.16: A sample set of the extracted hand-labeled occlusion edges.

Local statistics of hand-traced occlusion edges

Figure 2.17 show the distributions of Michelson contrast and RMS contrast for occlusion edges extracted using the hand tracings. In each sub-figure the horizontal axis shows the contrast values and the vertical axis shows the number of occlusion edge patches. Figure 2.17 (a) shows the distribution of Michelson contrast and Figure 2.17 (b) shows the distribution of RMS contrast. Figure 2.17 (a) demonstrates that the distribution of Michelson contrast is uniform with a bias

towards low contrast (see Discussion). Similarly, the RMS contrast distribution in Figure 2.17 (b) is roughly uniform with a bias towards low RMS contrast.

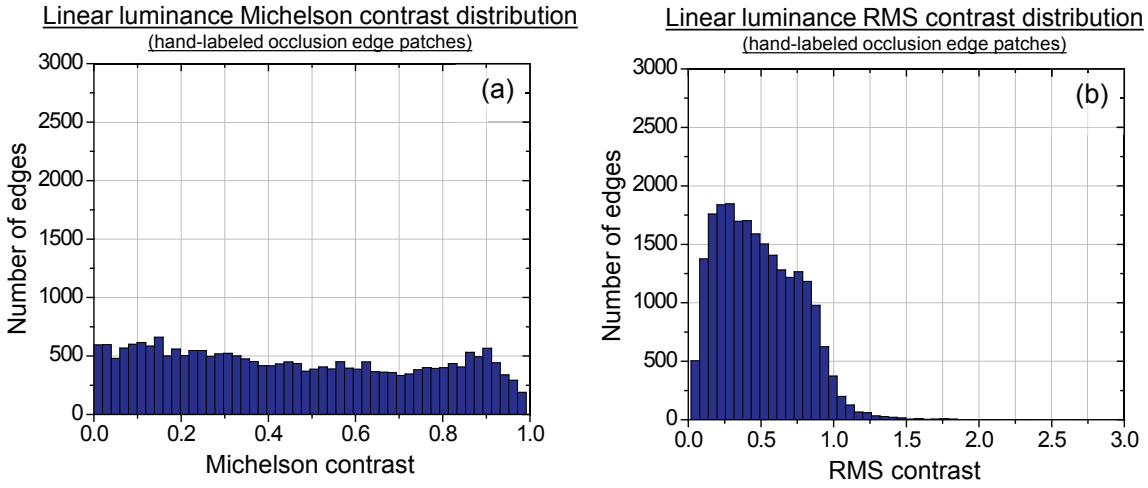


Figure 2.17: The distribution of hand-labeled occlusion edge contrast in natural scenes. (a) shows the distribution of Michelson contrast in hand-labeled occlusion edge patches. (b) shows the distribution of RMS contrast in hand-labeled occlusion edge patches.

Figure 2.18 (c) shows the one-dimensional normalized average occlusion edge for the hand-traced images. This plot is similar to the normalized average of the occlusion edges found by the Canny algorithm.

2.2.5 Edge classification using maximum likelihood classification

To investigate whether a local feature such as contrast can be used to classify an edge into occlusion or non-occlusion categories, we used the Michelson contrast as the local feature with a basic maximum likelihood classifier. The class which

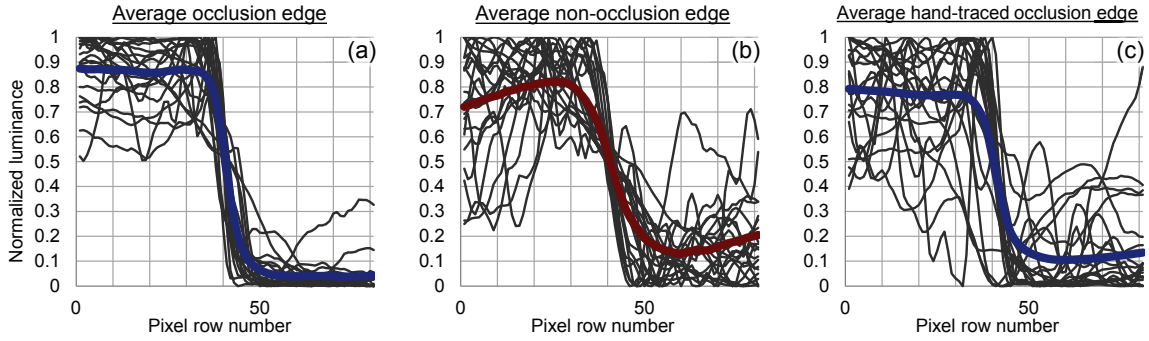


Figure 2.18: One dimensional profiles of normalized average occlusion, non-occlusion edges, and hand-traced occlusion edges. (a)The normalized average occlusion edge in blue with 20 sample occlusion edges. (b) The normalized average non-occlusion edge in red with 20 sample non-occlusion edges. The edges in (a) and (b) were first detected by the Canny operator and then categorized by participants as occlusion or non-occlusion edges. The slope of occlusion edges was significantly different ($t(671) = 16.08$, $p < 0.0001$) from the slope of non-occlusion edges. (c)The normalized average hand-traced occlusion edge in blue with 20 sample occlusion edges (details below in Human-labeled occlusion edges).

yields the maximum likelihood given the contrast of an unknown edge is the predicted class of that edge. The maximum likelihood for each edge category was computed using Bayes theorem:

$$\hat{Class} = \arg\max_{Class}(p(Class|Contrast)) = \arg\max_{Class}(p(Class) \times p(Contrast|Class)) \quad (2.6)$$

where, $\hat{Class}$ is the maximum likelihood estimate of an edge class, $p(Class|Contrast)$ is the probability of an edge class given edge contrast, $p(Class)$ is the prior probability of occurrence of an edge class and $p(Contrast|Class)$ is the probability of a contrast value given the class of an edge. 80% of the occlusion and non-occlusion edges were randomly selected and used to train the classifier and the remaining 20% of the occlusion and non-occlusion edges were used to test the classifier. This cross-validation scheme was iterated 100

times in order to measure the mean classification accuracy. All the occlusion and non-occlusion edges were first found by the Canny algorithm and extracted into edge patches as described before. The prior probability of an occlusion edge ($p(C = occlusion)$) was set to 0.4368 and probability of a non-occlusion edge ($p(C = non\text{-}occlusion)$) was set to 0.4338 based on 100% between-participant agreement. The remainder of the probability (0.1294) is accounted for by the edges which did not satisfy the 80% between-participant agreement ($p(C = indeterminate)$). The likelihood probability of contrast for each edge class was learned using the training edges.

Table 2.2 shows the prediction performance of the Michelson contrast as a local feature in classifying the remaining 20% of edges as occlusion and non-occlusion edges. Similar results were obtained using the RMS contrast as a local cue for prediction (not shown).

Table 2.2: The confusion matrix showing the prediction ability of the Michelson contrast as a local cue in predicting whether an edge is occlusion edge or non-occlusion edge.

		Predicted Class	
		Occlusion edge	**Non-occlusion edge**
True Class	**Occlusion edge**	**83.09%±4.41%**	16.91%±4.41%
	Non-occlusion edge	16.45%±4.53%	**83.55%±4.53%**

Figure 2.19 shows predictions for occlusion edges (in green) and non-occlusion edges (in red) edges in natural images, using the maximum likelihood classifier. To generate these predictions, the Canny edge detection algorithm was first applied to the original images to determine the edge locations. Then, for each edge location, an edge patch of 81×41 pixels was extracted. Finally,

for each extracted patch, the Michelson contrast was computed and used as the local feature for classification. For the first two images, most of the edges are correctly classified as occlusion and non-occlusion edges. However, the rightmost image suggests that our classifier could make numerous errors for some images. These results demonstrate that the contrast as a local feature is by itself a strong cue for predicting whether an edge is occluding or non-occluding.

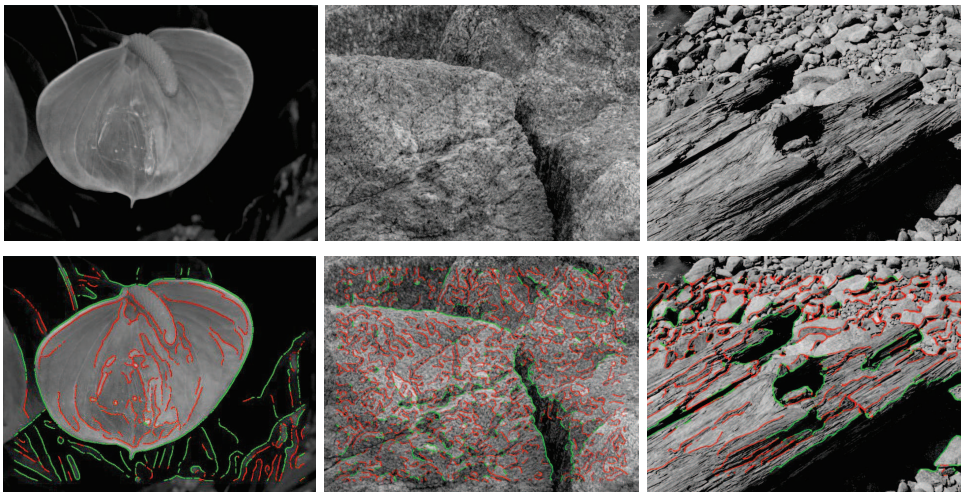


Figure 2.19: Predicted occlusion and non-occlusion edges using only contrast as local feature in the maximum likelihood classifier. The edges in green are the predicted occlusion edges and the edges in red are the predicted non-occlusion edges.

2.3 Discussion

In this study, we estimated the relative proportions of occlusion and nonocclusion edges. We examined the relative proportions of three subcategories of nonocclusion edges. We computed the statistics of local features (such as contrast and luminance) of occlusion and nonocclusion edges and built a classifier for unknown edges using the local information in the edge patch. The five main

findings of the study are as follows:

1. Given that an edge was detected by the Canny algorithm approximately half of the edges were labeled as occlusion edges and half as non-occlusion edges. There was good reliability across subjects, as only 5% of the edges did not satisfy the 80% between-participant agreement criterion.
2. When the edges are detected by a Canny operator, the average contrast of occlusion edges was found to be significantly higher ($P < 0.001$) than the contrast of non-occlusion edges.
3. A maximum-likelihood classifier with contrast as the only local feature could correctly predict 83% of human labeling decisions when classifying occlusion versus non-occlusion edges.
4. The contrast distribution of hand-labeled occlusion edges is approximately uniform with little bias towards low contrasts, whereas the distribution of occlusion edges found by the Canny algorithm is significantly different (2-sample Kolmogorov-Smirnov test, $D = 0.47$ $p < .001$). This implies that there are many occlusion edges that are easily identified by human observers, but will be missed by common edge detection algorithms such as Canny.
5. Non-occlusion edges due to cast shadows occur relatively rarely in our collection of natural scenes compared to surface-change and reflectance-change edges.

2.3.1 Occlusion edges versus nonocclusion edges

We found that 50% of edges detected by the Canny algorithm were labeled as occlusion edges by the human participants. We used 5 participants, and there was a good reliability across subjects. This was reflected as only 13% of the total edges in the experiment did not satisfy 100% mutual agreement (i.e., the agreement between 5 out of 5 subjects). In an unpublished study, Elder et al. (1999) performed a similar experiment of edge classification and discovered similar proportions of edge categories. We should note here that; these results depend on the kinds of image database used for the experiment. One could use a database of images where there are more occlusion edges than nonocclusion edges. Similarly, one could use texture images where there are more nonocclusion edges than occlusion edges. I will later discuss more about the limitations of the study.

One important finding of this study is that local feature statistics have significant information about the class of an edge. This implies that the early stages of the visual system could potentially start identifying the occlusion edges and thus start to separate figure from ground. Several object-recognition models take advantage of segregating figure from ground in initial steps of the algorithm (Leibe et al., 2008; Viola and Jones, 2001). Here we are not arguing that figure-ground segregation happens first in the visual system, but, that the local features extracted in the early stages assist in segregating figure from the ground. We believe that the visual system could begin to build probabilities about the edge classes from early stages of visual processing (e.g., the retina), and further refine these probabilities as more global features get available in the later stages of the visual processing.

There are some discrepancies in the plots showing the distribution of contrast (Figure 2.6a and 2.17a). Figure 2.6a shows the contrast distribution of the occlusion edges which were first detected by the Canny detector and then labeled as occlusion edges by human participants. It appears that most of the occlusion edges have high contrasts (75% of the occlusion edges have contrast more than 0.4) and very few have low contrasts. Similarly, Figure 2.17a shows the distribution of contrasts in occlusion edges identified by hand-tracing. However, in this figure, the distribution is relatively flat. We believe that the reason for this discrepancy is that for the hand traced occlusion edges the human observers can use a variety of long range cues to identify occlusions in an image. Observers can interpolate from far outside of the local area to estimate where an occluding edge may occur. Furthermore, a human observer can identify edges from texture boundaries that may be invisible to the Canny detector. We believe that the Canny detector often fails to detect low contrast occlusion edges; however, if the Canny detector finds a low contrast edge, then it is more likely to be a nonocclusion edge.

2.3.2 Nonocclusion subcategories

Nonocclusion edges were further classified into three subcategories (cast shadow edge, reflectance change edge, and surface change edge) by five participants. We designed a triangular slider for this classification task, where a participant could label an edge as combinations of three nonocclusion subcategories. The triangular slider was designed because there are several edges which have combinations of multiple edge types or do not have enough cues to confidently classify them as one of the edge types. The results indicated that

there are very few edges due to cast shadow. This could be either due to the lack of cast shadow edges in the image set we used for the experiment or the Canny detector missed cast shadow edges in the images because they were too spread out, blurred, or very low contrast. Most of the other nonocclusion edges were labeled as surface change and reflectance change edges. However, the mutual agreement across participants was low on the proportions of surface change and reflectance change edges. We believe that the cause of the low mutual agreement is due to the lack of cues while making the decision about the edge subclass. Also, the subcategories of nonocclusion edges are not mutually exclusive, and an edge could occur because of the co-occurrence of multiple causes of subcategories. Also, we do not believe that the variations in the proportions are due to a small number of participants or any misunderstanding of definitions of the subcategories. Figure 2.12 shows an example of nonocclusion edges which did not meet the 80% mutual agreement between participants. We can see that it is really a difficult task to make a definite decision about a subcategory. For example, when there is a crease on a rock or a vein on a leaf, it is difficult to determine whether there is a surface change edge, a reflectance change edge, or both.

2.3.3 Limitations of the study

Finally, I would like to discuss some of the limitations of this study. We used two approaches to locate edges in natural scenes. In the first approach, we identified edges using the standard Canny edge detection algorithm, which were then classified by human observers. In the second approach, we asked human observers to trace occlusion edges in natural scene images. Unlike the first ap-

proach, in the second approach, observers had full image information (global cues) to identify occlusion edges. Both the methods have limitations, and we believe there is no method which could provide ground truth classification of the causes of edges in natural scene images. The Canny edge detection algorithm and human observer both are limited by their biases. First, the Canny algorithm has parameters that can be varied. We believe the parameters we chose were reasonable and did not show any large differences in the results. However, there can be a different set of parameters that could potentially produce different results. Also, the Canny edge detection algorithms tend to miss on the edges which do not have significant luminance differences, whereas our results indicate that the hand-tracing approach could find edges which are very low in contrast. In the hand-tracing approach, observers could integrate over long range cues as well as use high-level knowledge of the object to identify edges. There are studies which also show that hand-labeled edges are not identifiable locally (DiMattina et al., 2012; McDermott, 2004). However, hand labeling has its own biases. For example, we cannot ask an observer to label each and every occluding edge in an image. We asked our participants to focus on the well-defined objects and not on the fine details (e.g., grass).

In this chapter, I presented the statistics of different classes of edges. I demonstrated that the local contrast contains the significant information to differentiate between occlusion and nonocclusion edges. The results suggest that there exists information regarding the cause of the edge at the earliest stages of the visual system where contrast can be estimated (i.e., the retina). The early visual system could potentially use this information to identify the causes of an edge. It is possible that neurons in the visual system which appear to be simple spot or edge detectors could be performing multiple computations to segregate

object from background.

CHAPTER 3

THE NONLINEARITIES IN THE VISUAL SYSTEM

A wide variety of studies have investigated the underlying processes of the mammalian visual system. The complexity of the visual system is due to its large number of neurons and the inherent nonlinearities at each stage of processing. The early efforts to understand the visual system tried to map the responses of neurons to simple stimuli. Single-cell recordings from the cat and the macaque visual neurons revealed the 2D spatial pattern (the receptive field) that described the neuron's response profile (e.g., Hartline et al. (1956); Hubel and Wiesel (1962)). Earlier models of the visual system used the linear systems approach to predict the response of a neuron to novel stimuli (e.g., Robson (1975); Shapley and Victor (1978); DeValois and DeValois (1988); Enroth-Cugell and Robson (1966); Movshon et al. (1978b)). The advantage of a linear model, is that it allows one to predict the neuron's behavior to any stimulus based on the response to some basis (e.g., spots or gratings). For example, response ($R(S)$) to a stimulus S can be computed as a simple dot product between the stimulus S and receptive field rf.

$$R(S) =< rf.S > \tag{3.1}$$

The receptive field as a complete description of a neuron is only valid if the neurons in the visual system are linear. However, biological neurons are highly nonlinear. The response of a biological neuron to a composite pattern is not a linear sum of the responses to the basis stimuli that compose the pattern (See Equation 3.2).

$$R(\sum_i x_i) \neq \sum_i R(x_i) \tag{3.2}$$

where x_i is i^{th} stimulus.

These early efforts based on linear systems approach were relatively successful in modeling the response behavior of the visual system to simple stimuli, but performed poorly when predicting the response behavior to complex stimuli such as natural scenes (Olshausen and Field, 2004; Carandini et al., 2005).

3.1 Nonlinearities in primary visual cortex

In 1950, Hubel and Wiesel were using spot stimuli to probe the receptive fields of the neurons in cat V1. They used spot stimuli because the retinal ganglion and LGN cells responded maximally to spot stimuli. However, Hubel and Wiesel accidentally found that neurons in V1 respond to elongated stimuli such as bars and edges at a variety of scales and orientations. They called these neurons 'V1 simple cells'. However, they also found some neurons that were tolerant to the position of edges within a 3-degree diameter of the receptive field (Hubel and Wiesel, 1962). This tolerance in the position was one of the first nonlinear effects observed in the neuron's response. They called these neurons 'V1 complex cells'. Hubel and Wiesel distinguished simple cells from complex cells based on following four properties:

1. V1 simple cells have distinct excitatory and inhibitory sub-regions within the receptive fields.
2. V1 simple cells responses are proportional to the linear summation within the subregions of the receptive field.
3. There is mutual antagonism between the excitatory and inhibitory subregions and balance out each other when stimulated with a uniform field stimulus.

4. Responses to novel stimuli can be predicted based on the arrangement of subregions.

Since Hubel and Wiesel (Hubel and Wiesel, 1962), a wide variety nonlinearities have been observed in visual cortex. Some of the nonlinearities in V1 are as follows:

1. **Nonlinear spatial summation in simple cells**: Movshon et al. (1978b) recorded from cat visual cortex using grating stimuli flickering at different locations. They found that most of the V1 simple cells respond according to the linear spatial summation as stated by the second property of simple cells described by Hubel and Wiesel (1962). However, a small number of V1 simple cells response was not modulated according to the linear spatial summation. Especially for high spatial frequency moving stimuli, these cells did not modulate according to linear summation.
2. **Contrast expansion and saturation (gain-control)**: Albrecht and Hamilton (1982) measured the responses of several neurons in V1 with grating stimuli of various contrast and spatial frequencies. They found that at lower contrasts (less than 6%) the response increased very rapidly. This part of the response is referred to as contrast expansion. A further increase in the contrast of the stimulus (above 6%) results in a response that increases linearly. However, because of the limited dynamic range, the response starts saturating at higher contrast. One important fact to be noted here is that the responses to different stimuli do not saturate at the same response magnitude. The stimulus with optimal spatial frequency, optimal orientation, and optimal spatial phase saturates the neuron and evokes its maximum firing rate, however the non-optimal stimulus (with non- op-

timal spatial frequency, non-optimal orientation, and non-optimal spatial phase) saturates at a lower response rate (Albrecht and Hamilton, 1982; Sclar and Freeman, 1982).

3. **Extra classical receptive field effects**: The first observation of the surround effect was made by Hubel and Wiesel (1965). They observed that as the bar length increased beyond a certain length, the neuron's response started decreasing. They called these neurons "hypercomplex" cells. This behavior is also referred to as end-stopping (Rose, 1977). Later, Cavanaugh et al. (2002) also measured the influence of the receptive field and its surround in V1 neurons.

4. **Cross-orientation inhibition**: Hubel and Wiesel (1962) found simple cells and complex cells in cat V1 selective to the orientation of the stimuli. However, they also observed a non-linear inhibitory process responsible for the orientation selectivity of the V1 neurons. Morrone et al. (1982) studied these non-linear inhibitory processes in V1 neurons to complex pattern stimuli. They found that the V1 neurons respond maximally to the optimally oriented grating and do not respond to stimuli orthogonal (non-optimal orientation) to the optimal stimulus. However, when the stimulus was generated by adding optimal and non-optimal stimuli, the response of the neurons reduced significantly.

5. **Spatial frequency inhibition**: Like cross-orientation inhibition, De Valois and Tootell (1983) found inhibition from non-optimal spatial frequency stimuli in the responses of cat V1 neurons. They measured neurons responses to optimal spatial frequency (f) gratings alone and superimposed with non-optimal spatial frequency gratings with frequencies of 1/4f, 1/3f, 1/2f, 2f, 3f, and 4f. Almost all of the simple cells showed reduced

firing when the optimal stimulus was superimposed with a non-optimal stimulus.

These are not the only nonlinearities encountered in the visual cortex. The vision science community has treated each nonlinearity as a technique employed by the visual system to solve some specific problem. Each nonlinearity has been modeled independently with separate mathematical equations and functional goals. The work that I will present in this chapter and the following chapters will attempt to describe a wide family of nonlinearities within a single geometric framework (some of this work has been published in Golden et al. (2016); Vilankar and Field (2017)). We will explore the geometry of neural responses and try to understand why different nonlinearities arise. We will focus on the inherent curvature of the iso-response surfaces of the neurons of various models. I will demonstrate how this curvature could describe a wide family of nonlinearities. We will compare various models that produce this curvature. We will focus on the sparse coding network model (Olshausen and Field, 1996) and understand the principles behind the curvature. We will further analyze how the different learning rules of sparse coding network affect the neural response geometry and how it affects the objectives of the network. And finally, I will explore how curvature produces hyperselectivity in visual neurons and breaks the Gabor-Heisenberg limit and its implications.

3.2 The image state space

The image state space is the geometrical space where we can represent the response characteristics of neurons (by plotting iso-response curves). We will ex-

plore how the response characteristics are different or similar for the types of nonlinearities and how it is affected by the parameters of different models. Image state space is the high-dimensional space where each dimension represents a pixel intensity. For example, the set of all possible images composed of 100 pixels is represented by a state space that is 100-dimensional, where each dimension represents a pixel intensity. In this space, each point represents an image. However, for the purpose of visualization, we will use lower dimensional image state spaces (two- and three-dimensions) and two-dimensional subspaces from the higher dimensional image state space. We believe that understanding the neural response behavior in low dimensions provide considerable insights into the efficient mechanisms employed by the visual system.

3.2.1 A linear neuron in image state space

If a neuron is linear, then it can be represented as a vector in image state space. The direction of the vector would also represent the optimal stimulus for that neuron. V1 is often modeled as an array of neurons (e.g., wavelet or Gabor functions) represented as vectors that span the image state space. Figure 3.1 shows an example of a linear neuron in two-dimensional state space. If the response of a neuron depends linearly on only the intensity of pixel 1, then it can be represented as a vector pointing in the direction of Dimension1 (D1) or pixel 1 (see Figure 3.1a). The response of this neuron would increase linearly with the increase in pixel 1 intensity. Figure 3.1b shows the response manifold of this linear neuron on the z-axis. A linear neuron can produce both positive and negative responses. In the example shown in the figure, the response of the neuron depends on the intensity of pixel 1 (D1). As D1 increases, the response

also increases. Figure 3.1a also plots the iso-response lines for the neuron. For a linear neuron, iso-response lines are always orthogonal to the direction of the vector representing the neuron. One more important characteristic of a linear neuron is that the iso-response lines are equally spaced in the image state space. In n-dimensional image state space, the iso-response surface of a linear neuron will be an n-1 dimensional surface orthogonal to the vector representing the neuron (i.e. the basis function such as a Gabor or wavelet).

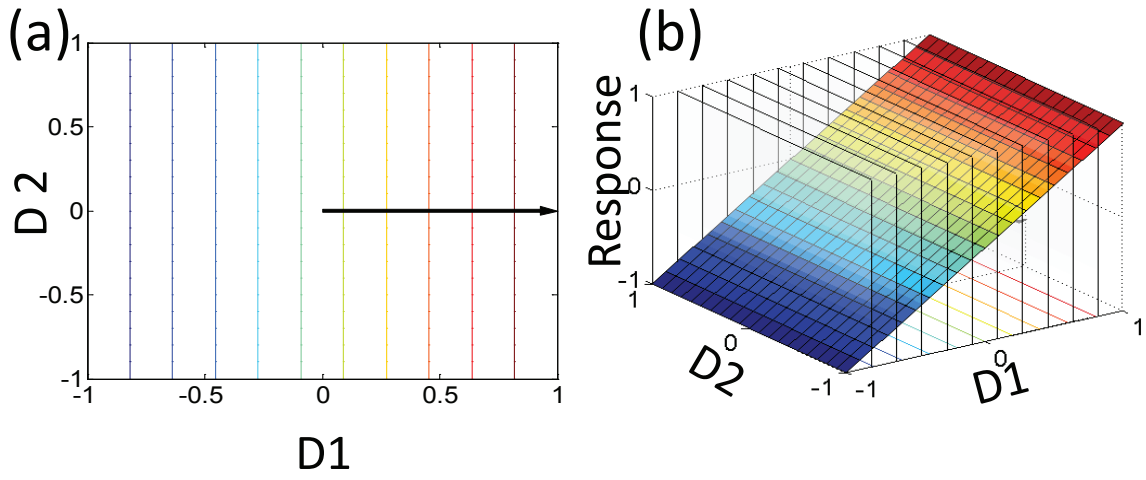


Figure 3.1: The figure shows the response geometry of a linear neuron in a 2-dimensional image state space. In (a) the neuron is represented as a vector [1,0]. Each colored orthogonal line is an iso-response contour which represents a set of stimuli in the image state space. (b) shows the response surface where the Z-axis represents response magnitude of the neuron.

3.2.2 Thresholded non-linear neuron

The first and the simplest form of nonlinearity is the thresholded nonlinearity. In this form of nonlinearity, a neuron responds only if the response is above some threshold magnitude. For example, a thresholded neuron responds pos-

itively and responds with zeros if the response to a stimulus is negative. This kind of nonlinearity is also referred as an output or point-wise nonlinearity. Figure 3.2b shows the response manifold for a thresholded nonlinear neuron. The response is zero for the negative part of the stimulus (pixel 1 intensity is below zero), and when the stimulus intensity exceeds zero, the neuron begins to respond linearly. All biological neurons have this kind of nonlinearity as they cannot respond negatively.

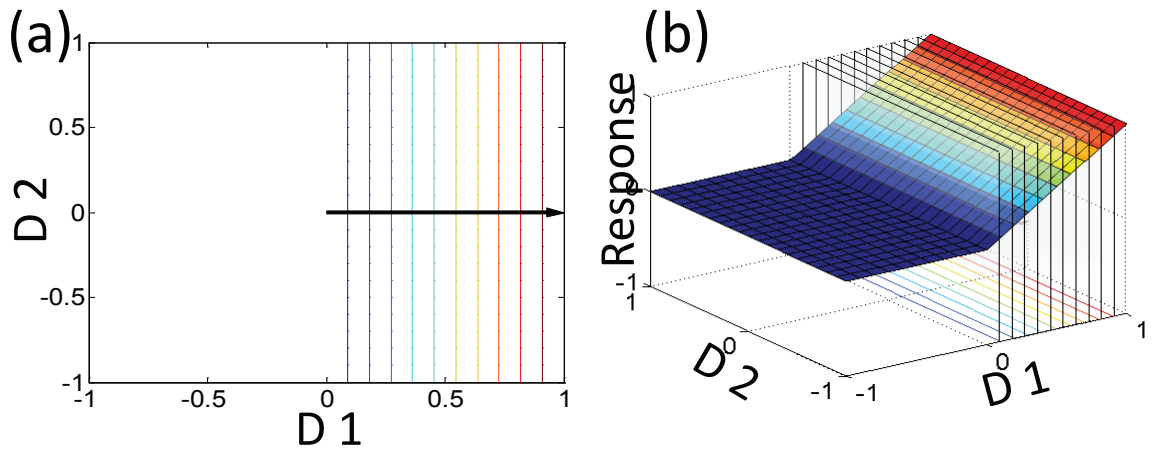


Figure 3.2: The figure shows the response geometry of a thresholded linear neuron in a 2-dimensional image state space. (a) shows the iso-response contours and (b) shows the response magnitude surface.

3.2.3 Compressive nonlinearity

Many neurons in the brain exhibit a compressive nonlinearity. In this kind of nonlinearity, the rate of increase in the response slows down with the increase in the stimulus intensity. This nonlinearity is also an output nonlinearity because an output nonlinearity is applied to the linear response of a neuron. Figure 3.3b shows the compressive nonlinear response manifold. The important character-

istic of this nonlinearity is that the iso-response lines (see Figure 3.3a) are still orthogonal to the vector. However, the spacing between the iso-response lines of different magnitudes changes with the stimulus intensity.

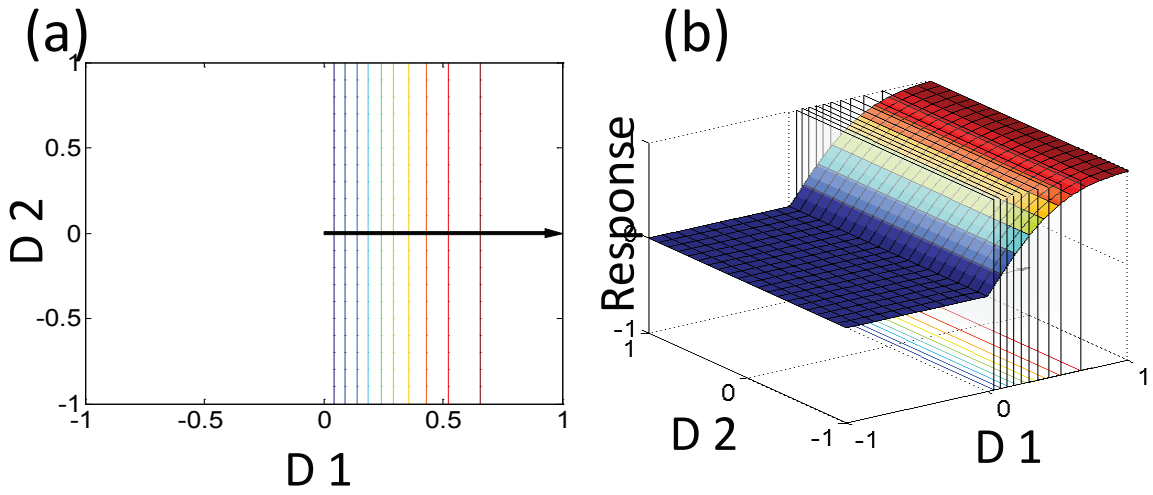


Figure 3.3: The figure shows the response geometry of a compressive non-linear neuron in a 2-dimensional image state space. (a) shows the iso-response contours, one should note that the iso-response contours are straight and orthogonal to the vector. (b) shows the response magnitude surface.

3.2.4 Warping nonlinearity

The nonlinearities discussed so far are output nonlinearities, where the nonlinearity (e.g., threshold, compression, and expansion) was applied to the linear output of a neuron. These nonlinearities are also referred as planar nonlinearities because the iso-response lines or surfaces (in high dimensions) are straight lines or planes and perpendicular to the direction of the vector. These nonlinearities can be observed in V1 neurons. However, they fail to explain many nonlinear behaviors that we discussed before (e.g., end-stoping, gain control,

etc.). Many modeling efforts have demonstrated that planar nonlinearities are not sufficient to capture a variety of complex behavior of neurons in the early visual system (e.g., Albrecht and Geisler (1991); Heeger (1992); Tolhurst and Heeger (1997)).

Here, I would like to describe a new form nonlinearity (introduced in Golden et al. (2016)) called a warping nonlinearity. This warping nonlinearity could describe a wide family of nonlinear behavior of visual neurons. Figure 3.4a shows a neuron's iso-response contours with a warping nonlinearity. In this kind of nonlinearity, the iso-response lines are not straight and perpedicular to the direction of the vector. They get warped either away or towards the origin, and this simple warping of the iso-response contours is responsible for the complex nonlinear behavior of the neurons. Zetzsche et al. (1999) have also demonstrated that iso-response lines that bend away from the origin produce a family of nonlinearities observed in V1.

3.3 Exo-origin and endo-origin curvature

Depending on how the iso-response contours are curved, we define two forms of curvature (Golden et al., 2016). Figure 3.5a shows the iso-response contours which bend away from the origin. This curvature in iso-response contours is referred as exo-origin curvature. We define exo-origin curvature as a curved iso-response line where there exist no two points on the curve such that the line segment connecting the two points passes through the origin. This form of curvature in the iso-response contours of a neuron produces a family of nonlinearities found in V1. For example, exo-origin curvature produces end-stopping,

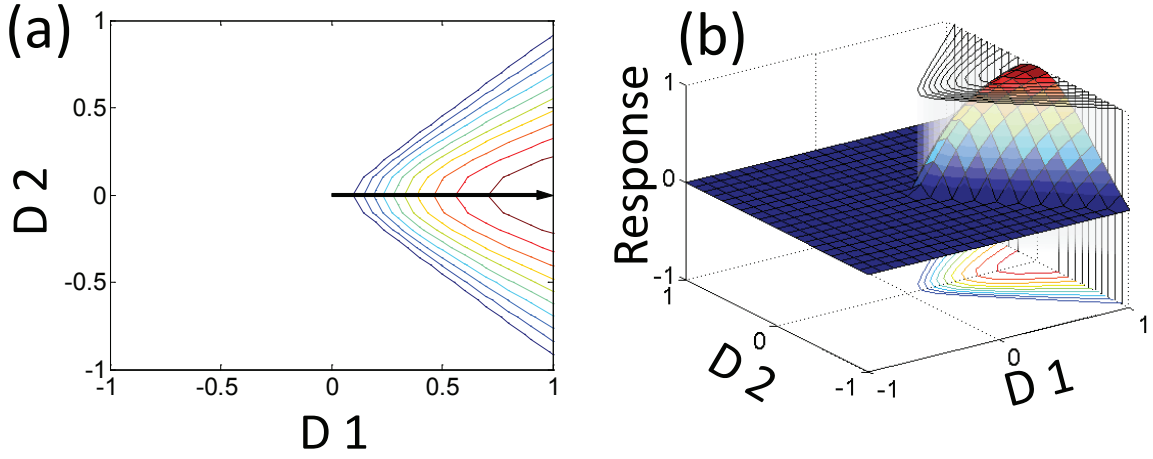


Figure 3.4: The figure shows the response geometry of a warping nonlinear neuron in a 2-dimensional image state space. (a) shows the iso-response contours, one should note that the iso-response contours are curved and warped around the vector.(b) shows the response magnitude surface.

nonclassical receptive field effect, gain control, etc. The exo-origin curvature causes a neuron to respond to a smaller region of the image state space compared to a linear neuron and hence such neurons become hyperselective. The exo-origin curvature makes a neuron's response hyperselective to some stimulus feature (e,g., size, shape, contrast). Hence they are also referred as selective nonlinearities. I will demonstrate in the next chapter, how the exo-origin curvature could explain the selective nonlinearities. Zetzsche et al. (1999) have also demonstrated that a simple curvature in the iso-response lines away from the origin could explain many nonlinearities observed in V1.

Figure 3.5b shows the iso-response contours which bend towards the origin. This curvature in iso-response contours is referred as endo-origin curvature. We define endo-origin curvature as a curved iso-response line where there exist at least two points on the curve such that the line segment connecting the two

points passes through the origin. The iso-response contours with endo-origin curvature produce invariant nonlinearities (e.g., complex cells). The invariant nonlinearities are the nonlinearities where a neuron is invariant to some stimulus feature (e.g., phase, position, etc.).

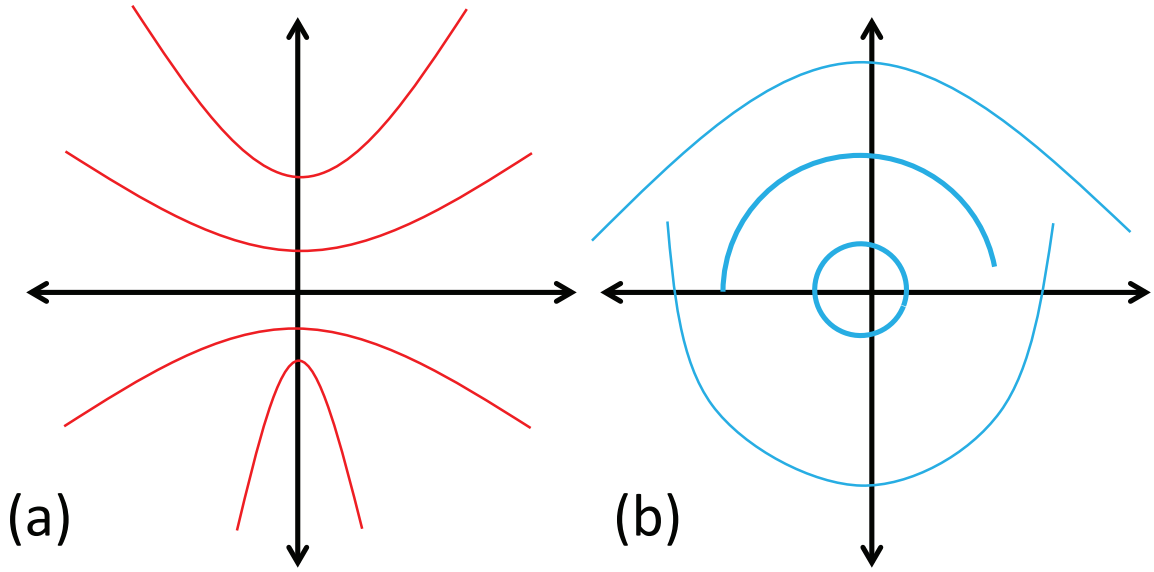


Figure 3.5: Examples of proposed curvatures in the iso-response contours. a) Shows examples of exo-origin curvature (curved away from the origin) and b) shows examples of endo-origin curvature (curved towards the origin).

In the next chapter we will explore how the exo-origin and the endo-origin curvature describes a wide family of nonlinearities observed in V1.

CHAPTER 4

A GEOMETRICAL PERSPECTIVE OF NONLINEARITIES

For linear neurons or neurons with planar nonlinearities, the iso-response contours do not warp, and they remain perpendicular to the vector representing the neuron. In the last chapter, I introduced two forms of warping in the iso-response contours, where the iso-response contours are not straight but warp either away from the origin or towards the origin. A neuron with exo-origin curvature is the neuron which has iso-response contours curved away from the origin and a neuron with endo-origin curvature is the neuron which has iso-response contours curved towards the origin. In this chapter, I will show how these curvatures in the iso-response contours can describe a wide family of nonlinearities observed in V1. I will demonstrate how exo-origin curvature can describe selective nonlinearities such as extra-classical receptive field effects and how endo-origin curvature can describe invariant non-linearities such as a complex cell (phase invariant/tolerant neuron).

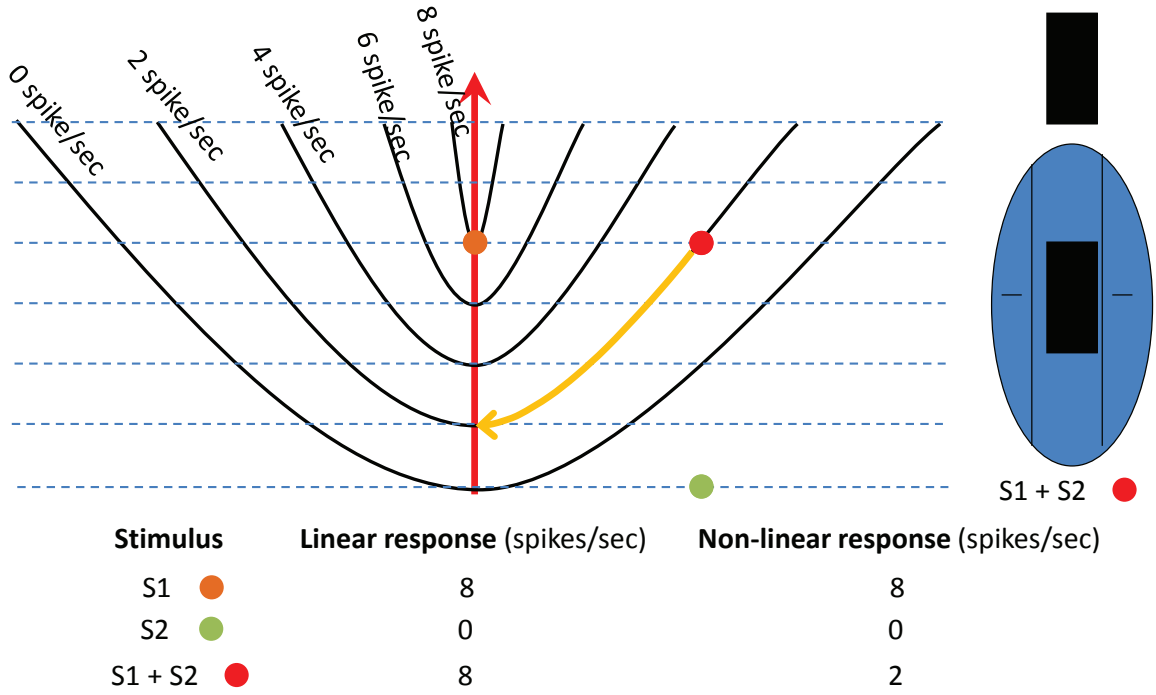
4.1 Exo-origin curvature and non-classical receptive field effects

As first noted by Hubel and Wiesel (1965), the presence of a stimulus in certain locations outside the classical receptive field region can reduce the firing of the neuron. They found neurons where the firing increased with the increasing length of a bar stimulus inside the classical receptive field. However, when the length of the bar spilled beyond the classical receptive field and into the surround, the firing rate decreased. This phenomenon was also later discovered

by Rose (1977). They found that half of the cells in cats visual cortex produce weaker responses to the long bar stimuli than to the short bars. They named this phenomenon end-stopping. Later, Cavanaugh et al. (2002) also discovered this phenomenon in macaque V1 cells with grating stimuli. We believe that looking at the geometry of exo-origin curvature can describe this phenomenon. Here I will demonstrate this with a toy example in a two-dimensional subspace of high dimensions.

Figure 4.1 shows a red vector representing a neuron. This vector is pointing in the direction of the neurons preferred stimulus (a bar) represented by the green point (A). We will consider two scenarios. In one scenario, we will assume the neuron is linear and hence it will have iso-response contours straight and perpendicular to the vector (iso-response contours shown as blue-dashed lines). In the second scenario, we will assume the neuron has exo-origin curvature in its iso-response contours (iso-response contours shown as black-solid lines). In both the scenarios, the linear and nonlinear neuron will respond with 8 spike/sec for the optimal stimulus (green point). For this neuron, any stimulus in its non-classical receptive field will be orthogonal to the direction of the vector representing the neuron (represented by the gray point B). Again, in both the scenarios, the linear and the nonlinear neuron will not respond (0 spikes/sec). Next, we probe this neuron with a hybrid stimulus A+B. This stimulus could be a long bar which has part of the bar inside the classical receptive field and the ends of the bar in the non-classical receptive field. Adding stimulus B to stimulus A will place the hybrid stimulus at that point in the image state space represented by the red dot. For the scenario where the neuron is linear, this addition of stimulus B to A will not affect the firing rate (the response will be 8 spike/sec according to the iso-response line passing through the red point).

However, in the second scenario where the neuron has exo-origin curvature in its iso-response contours, the response will be reduced to 2 spikes/sec. This is because the point representing the hybrid stimulus now intersects with the iso-response contour of 2 spike/sec.



Stimulus	Linear response (spikes/sec)	Non-linear response (spikes/sec)
S1	8	8
S2	0	0
S1 + S2	8	2

Figure 4.1: The figure shows Exo-origin curvature and how it describes non-classical effects like end-stopping, cross-orientation inhibition, etc. In this example Stimulus 'A' represents the most effective stimulus for the neuron. For the magnitude shown, stimulus 'A' elicits 8 spikes/sec in the neuron. For instance, this could represent a bar presented in the center of the neurons receptive field at its preferred orientation. Stimulus 'B' represents a stimulus that produces no response in the neuron. For example this could represent a bar presented outside the classical receptive field. Although 'B' (or 'B') produces no response on its own, when stimulus 'B' is combined with stimulus 'A', the neurons response will be reduced. Both end-stopping and cross-orientation inhibition are examples of this general form of non-linearity.

4.2 Exo-origin curvature and gain-control

Exo-origin curvature can also describe the gain control behavior of neurons in V1. Figure 4.2 shows the response behavior of a neuron to a variety of grating stimuli (Albrecht and Hamilton, 1982; Albrecht et al., 2002, 2003). The figure shows the response of a neuron as a function of grating stimuli contrast. Each curve in the figure shows the response for the grating stimulus of a particular spatial frequency. Some of the gratings produced a higher response (see the curve with solid circles) while other gratings produced a lower response (the curve with triangles). There are two important points to take from this gain control behavior. One, the response of the neuron increases and then saturates at higher contrast for all grating stimuli. Secondly, the contrast at which the response saturates was roughly the same. Albrecht et al. (2003) fit these curves using the Naka-Rushton equation (see Equation 4.2). The grating stimuli used for this experiment were taken from an orthogonal basis set. If we assume that the Naka-Ruston equation also holds for the stimuli between any two orthogonal stimuli, then we can create the response surface between the orthogonal stimuli such that the neuron saturates roughly at the same contrast for all the intermediate stimuli (response surface shown in Figure 4.3). One can see that this produces exo-origin curvature in the response profile (each color on the response surface represents the magnitude of the response). The response surface in Figure 4.3 was generated using the following equations. Note that the response surface is three-dimensional where the x-axis and the y-axis correspond to the two orthogonal grating stimuli, and the z-axis corresponds to the response of the neuron (this neuron has its optimal stimulus pointing along the x-axis).

$$resp = f_{NR}(r, n, r_{half}, V_{max}) \times \exp \frac{-\theta^2}{2 \times \sigma^2}, \tag{4.1}$$

$$f_{NR}(r, n, r_{half}, V_{max}) = \frac{V_{max} \times r^n}{r^n + {r_{half}}^n} \tag{4.2}$$

f_{NR} is the Naka Rushton equation with four parameters, V_{max} is the saturating value, r_{half} is the half saturation level, n is the exponent (for Figure 4.3 a,b and c n is set to 4), and r is the contrast or radial distance of a stimulus from the origin. $resp$ is the gain-controlled response, modeled as f_{NR} multiplied by a radial Gaussian with a 0 degree mean and 30 degrees of σ, θ is the polar angle of a stimulus in a 2D state space.

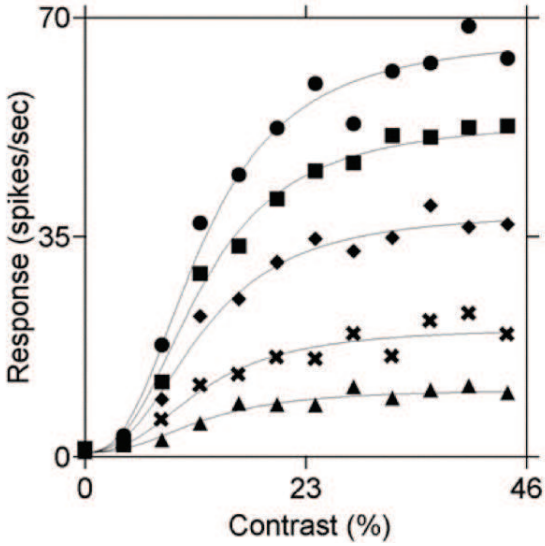


Figure 4.2: The figure shows response of a neuron that saturates at different response magnitudes for different stimuli (gratings of different spatial frequencies) but saturates at roughly the same stimulus magnitude (contrast).The figure is taken from Albrecht et al. (2003)

If we consider an array of stimuli shown as rays on the figure, then these stimuli will trace the response profile (see Figure 4.3b and c) similar to the response profile discovered by Albrecht and Hamilton (1982) in V1 neurons. They found that majority of neurons (70% of neurons) show this behavior. The main point of this figure is that this general nonlinearity caused by warping of the

iso-response contours can also produce gain control like behavior.

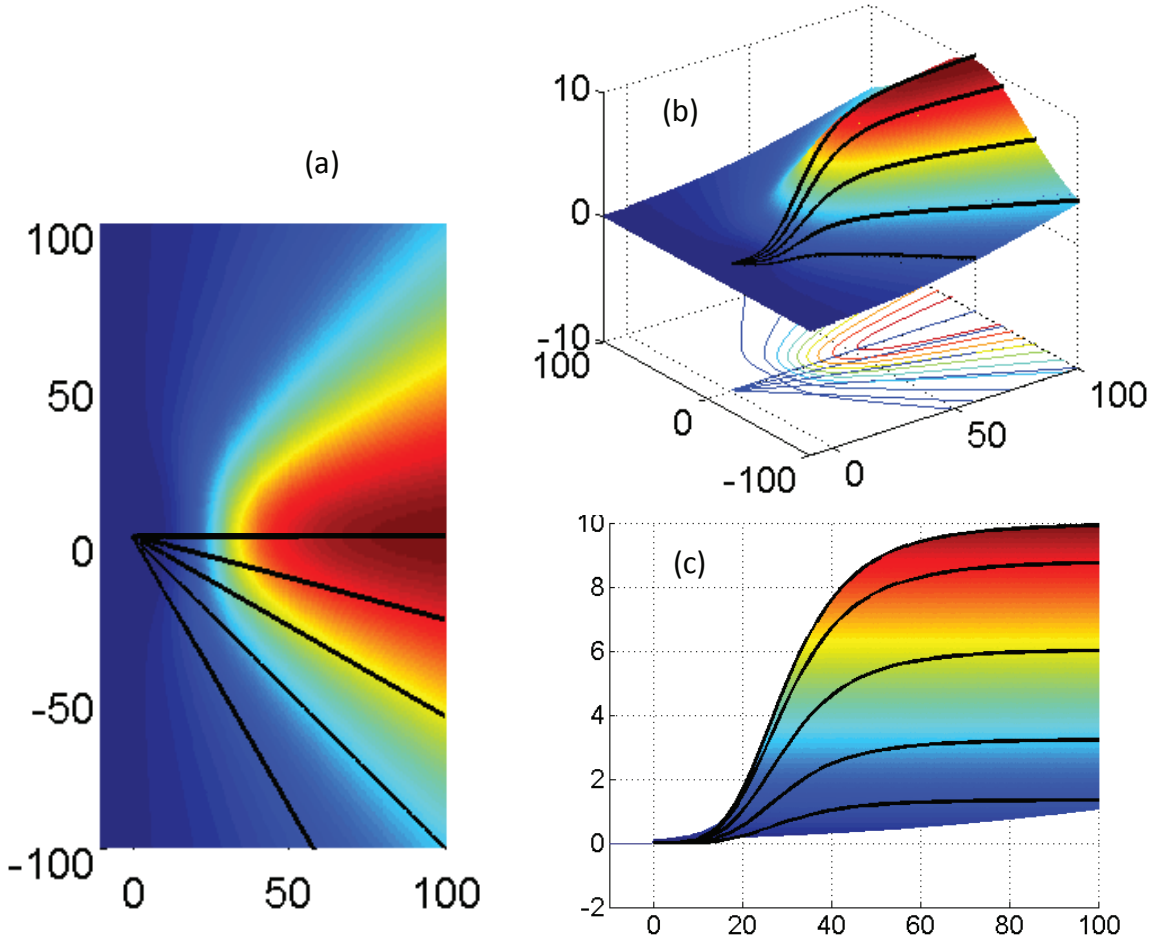


Figure 4.3: The geometry of gain control. Figure 4.2 shows the response to an orthonormal basis (sinusoids).If we assume that the neuron saturates at the same contrast for all stimuli between the orthonormal basis, we can generate a response manifold. (a) shows the response manifold computed using Equation 4.2. Here we are assuming that Equation 4.2 describes the contrast response for all stimuli. The black rays extending from the origin represents a particular stimulus of varying contrast. (b) shows a side view of this response surface along with the iso-response contours of the neuron and d) shows the contrast response generated with this response surface, where contrast is defined as the distance of a point from the origin. The intention of this figure is not to accurately model the gain-control behavior. Rather the intention is to demonstrate the relation between the geometry and the contrast response.

Many vision scientists have modeled gain control behavior. The standard model for gain control is divisive normalization (Heeger, 1992). In this model, the activity of each neuron is divided by the sum of activities of neighboring neurons. The simplest form of this model assuming only two neurons in the neighborhood is

$$resp = \frac{r_1}{\frac{r_1+r_2}{2} + 1} \tag{4.3}$$

where r_1 and r_2 are the squared linear responses of two orthogonal neurons. Figure 4.4a shows the side view of the response surface computed using Equation 4.3 and Figure 4.4b shows the top-view of the iso-response contours for the corresponding response surface in Figure 4.4a. The black lines extending from the origin shows the stimuli of varying contrast. From the iso-response contours from Figure 4.4b, we can see that the neuron's response computed using the divisive normalization equation produces exo-origin curvature. Also, the black curves representing the response to stimuli of increasing contrast on Figure 4.4a show a response profile similar to that of the gain control neuron in Figure 4.2. Other models of gain control (e.g., the fan equation model, see below) also show

similar exo-origin curvature in iso-response lines. For a detailed comparison of different models, please refer to Golden et al. (2016). One should note that we are not arguing in support of any particular model or providing evidence for the best model of gain-control. Currently, we do not believe there are sufficient physiological data available to distinguish between these models. Here we want to emphasize that these different models of gain-control also produce exo-origin curvature.

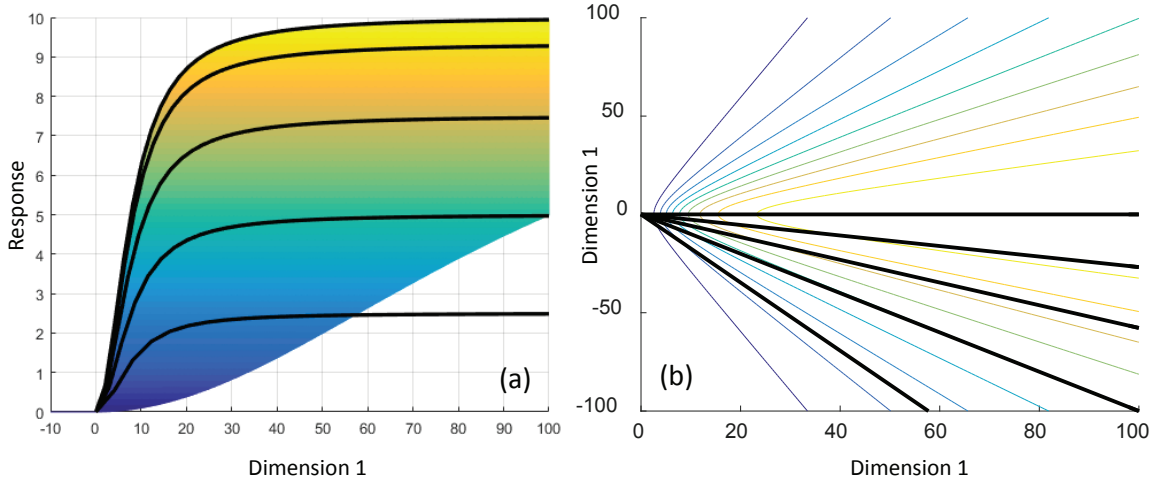


Figure 4.4: The figure shows the geometry of the divisive normalization model in 2D (Equation 4.3). (a) shows the response manifold surface, and (b) shows the iso-response contour. The rays extending from the origin represent stimuli of varying contrast. The stimuli radially distant from the origin have comparatively higher contrast from the stimuli near the origin.

Later in this chapter, we will discuss a variety of models which can produce exo-origin curvature. We will focus on the sparse coding network model and will discuss the principle behind producing curvature in iso-response contours. We will demonstrate that a single straightforward objective of the network to efficiently represent natural scenes will produce exo-origin curvature which is responsible for non-linearities in the visual system which appear to be different

from each other.

4.3 Endo-origin curvature and invariant/tolerant nonlinearity

In macaque V1, it is estimated that 3/4 of the neurons show invariant/tolerant nonlinear behavior (Kagan et al., 2002). These neurons are referred as complex cells. Complex cells show little or no modulation in response to drifting grating stimuli. The response of a complex cell is modeled as the squared sum of even and odd-symmetric simple cells. This model was first popularized by Adelson and Bergen (1985) as the energy model and was a good first approximation to complex cell behavior. However, the model failed to predict response to natural scene stimuli (Prenger et al., 2004; Touryan et al., 2005). Figure 4.5a plots the response surface of the energy model. The response profile is a three-dimensional plot, where the x-axis and the y-axis represent the response of odd and even symmetric simple cells. These two simple cells have optimal stimulus that are orthogonal in image state space. The iso-response contours are perfect circles (endo-origin curvature) for this model which will produce an invariant response profile to the varying phase of the stimulus (shown in Figure 4.5b).

However, neurons in V1 are not all simple cells with feature selectivity (selectivity to shape, orientation and spatial frequency) and complex cells with complete invariance (e.g., Dean and Tolhurst (1983); Skottun et al. (1991)). There is a continuous distribution of cells between these two extremes, with varying degrees of tolerance to phase. An example of such a cell is shown in Figure 4.5c, where the iso-response contours bend towards the origin but do not close as they do for perfect complex cells. Such neurons show some tolerance to the

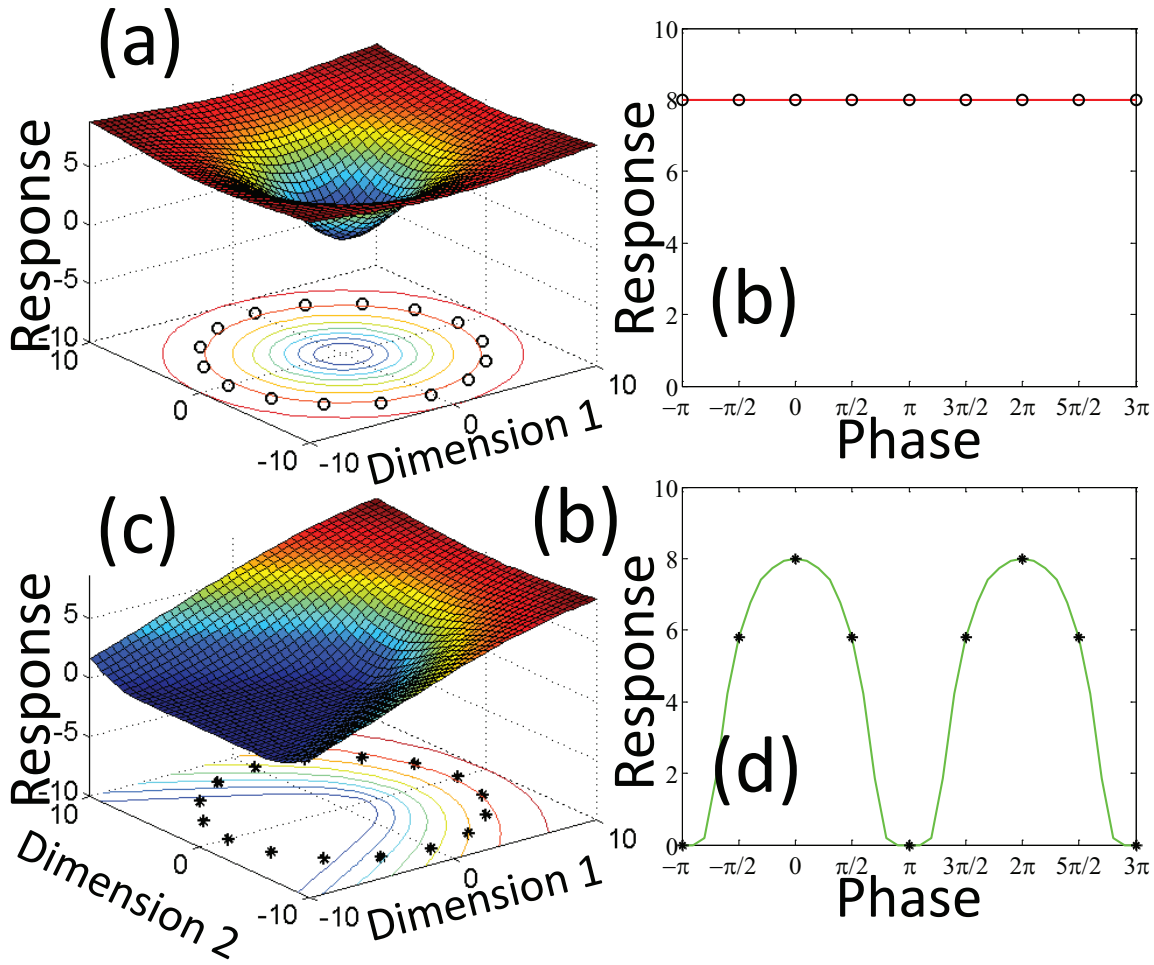


Figure 4.5: The figure shows endo-origin curvature. Here a V1 neurons response is modeled to a drifting sinusoidal grating using four models of endo-origin curvature. (a) represents the complex cell (energy model) which has perfect circulur iso-response contours. (b) shows the flat response of the complex cell as a function of the phase. (c) and (d) represent models of neurons that bridge the range between simple and complex cells. V1 neurons show a range of behavior between simple and complex (e.g., Dean and Tolhurst (1983))

phase of the stimuli, but it will oscillate as a function of the phase (shown in Figure 4.5d).

Currently, no single layer network model can produce endo-origin curvature

in its iso-response contours. There are multi-layer networks which can produce the endo-origin curvature. Also, there are models which can produce exo-origin and endo-origin curvature simultaneously. The analysis of endo-origin curvature is going to be investigated in future works and is currently beyond the scope of this dissertation (see Golden et al. (2016)) for a discussion on the principles behind the simultaneous exo-origin and endo-origin curvature).

4.4 Models with exo-origin curvature in two-dimensional image state space

So far, we have seen that simple curvature away from the origin in the iso-response contours can produce a wide family of nonlinearities observed in V1 neurons. However, now we will explore some of the models that generate this curvature without explicitly modeling the nonlinearities of V1. Four models that produce this curvature are:

1. Sparse coding (e.g.,Olshausen and Field (1996))
2. Fan equation (Golden et al., 2016)
3. Gain control with divisive normalization (e.g., Heeger (1992); Schwartz and Simoncelli (2001))
4. Cascaded linear-nonlinear model (Pagan et al., 2016)

Figure 4.6 shows the two-dimensional exo-origin curvature produced by the four modeling approaches mentioned above. Each of these models has multiple parameters which could affect the magnitude of the curvature of the iso-

response contours. Here we will compare these models by observing the curvature and how it is affected by neighboring neurons. Each row of the figure respectively represents the curvature produced by the models mentioned. The first column shows an example of a neuron and its iso-response contours produced by the model. The second column represents how the iso-response contours interact when the neighboring neurons are orthogonal. That is the vectors representing the neurons are orthogonal in image state space. The third column shows the interaction of the iso-response contours when the neighboring neurons are not orthogonal (the vectors representing the neurons are 60 degrees apart in image state space).

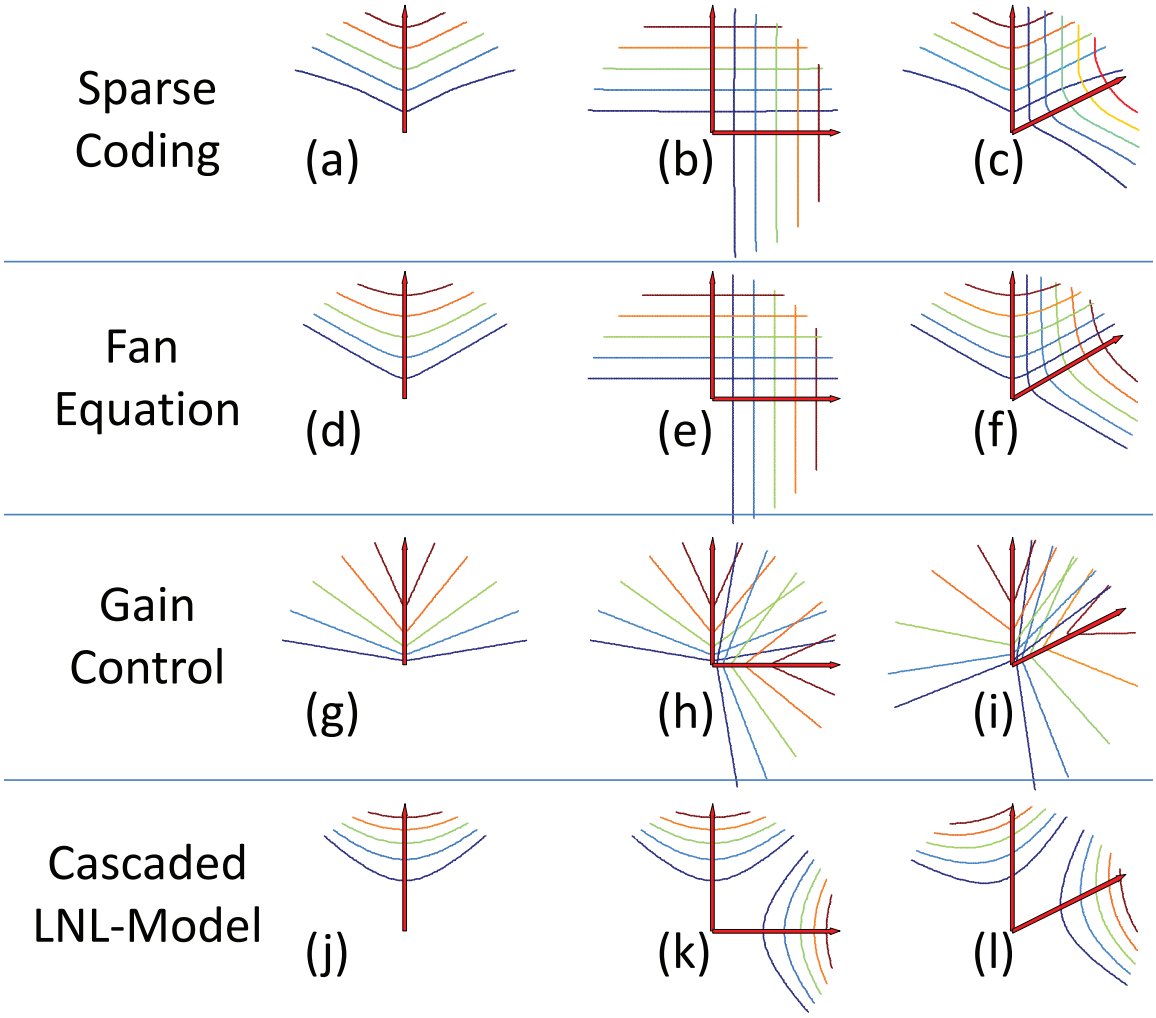


Figure 4.6: The figure shows the types of curvature produced by four models of V1 nonlinearities. Each of these approaches can produce hyperselctivity of variable magnitude. For each model we plot the iso-response contours in two-dimensions. We show these contours for a single neuron and show the contours for two neurons when the neurons are either orthogonal(second column of the figure) or not orthogonal(third column). The four approaches are 1)Sparse coding (a),b),c))), 2)Fan Equation (d), e), f)), 3)Gain control (g), h), i)), and 4) Cascaded linear-non-linear model (j,k,l). For sparse coding and the Fan equation models, the curvature depends on the angle between neighboring neurons(angle in image state space). If the neighbors are orthogonal, there is likely to be no or little curvature. For gain control, the curvature depends on whether the neighboring neuron is part of the group involved in divisive normalization. This can produce curvature even in cases where the neurons are orthogonal. As one can see, each of these approaches curves the iso-response contours differently. More critically, the grid of the iso-response contours will cover image space in different ways for each of these models.

The first row (Figure 4.6a,b,& c) represents the curvature produced by sparse coding network. The sparse coding network (Olshausen and Field, 1996) is a neural network which learns efficient representations of natural scene data. The bases learned with the sparse coding network resemble the receptive fields observed in V1. In the sparse coding network emphasis is typically given to the receptive fields, but here we focus on the response geometry in two-dimensional image state space. We can see that the sparse coding network produces no curvature when the neighboring angles are orthogonal (Figure 4.6b). In this case, the neurons behave linearly. However, when neurons are only 60 degrees apart, the network produces curvature in the iso-response contours. In the next chapter, we will explore the sparse coding network in more details. We will quantify the curvature produced by the sparse coding network in higher dimensions and examine how the different parameters of the network affect the curvature and

hyperselectivity of a neuron.

In Golden et al. (2016) we found a good approximate model for the curvature of response surfaces generated by the sparse coding network. In this model, the curvature is a straightforward function of the angle between the neighboring neurons in two-dimensional image state space. We call this model the fan equation model. The equation of the model is as follows:

$$\begin{aligned} a_i = f_{Fan}(c, \theta) &= c \times \cos(n(f_i, f_j)\theta) \\ n(f_i, f_j) &= \frac{\pi/2}{\arccos(\frac{\langle f_i, f_j \rangle}{\|f_i\| \|f_j\|})} \end{aligned} \tag{4.4}$$

where c is the distance of a stimulus from the origin (i.e. the stimulus contrast), θ is the angle between a stimulus and the neuron, a_i is the response magnitude of a neuron i, n determines the curvature and f_i and f_j are the vectors representing the two neurons. n is a function of the angle between neighboring neurons. When $n = 1$ the iso-response contours are flat (e.g., linear); when $n > 1$ the neuron has iso-response contours with exo-origin curvature and when $n < 1$ the neuron has iso-response contours with endo-origin curvature.

The curvature from the sparse coding network and the fan equation are very similar. In both, the model's curvature is maximum near the vector representing the neuron. Away from the vector, the curvature tends to flatten out. The amount of curvature also depends on the neighboring vector angle. For orthogonal neighbors, we see little or no curvature. However, curvature increases as the angle between the vectors representing the neighboring neurons decreases. In the next chapter, we will explore in more detail the functional relationship between angle and the curvature in high-dimensional sparse coding network.

Figure 4.6g,h and I show the curvature produced by the divisive normaliza-

tion gain control model (Equation 4.3). Unlike the sparse coding model and the fan equation model, the gain control model seems to produce curvature even when the neighboring neurons are orthogonal. The division from the activities of neighboring neurons makes a neuron even more hyperselective. One important thing to note which is different from the aforementioned models is that the curvature (hyperselective) increases with the stimulus magnitude.

Figure 4.6j,k, & l is generated by a cascaded linear-nonlinear model (Pagan et al., 2016). The network in this model usually consists of two layers of linear-nonlinear stages. In the first layer, the linear output (r_1 and r_2) is computed as the weighted sum of the inputs. The output is then sent through a squaring nonlinearity. At this stage, there is no curvature produced in the iso-response contours. The squaring nonlinearity is just a planar nonlinearity which does not curve the iso-response contours. The second layer then linearly combines the output of the previous layer to get the output. The output from the second stage produces the curvature in the iso-response contours thus increasing the hyperselectivity. The output from such a network can generally be written as the following equation:

$$resp = ar_1^2 + br_2^2 + cr_1r_2 + dr_1 + er_2 + f \tag{4.5}$$

where r_1 and r_2 are the linear outputs from the first layer. This equation can generate a wide variety of curvature in the iso-response contours from hyperbolas to ellipses in lower dimensions. For a more restricted family of quadratic curves used to plot the Figure 4.6j,k & l can be represented as:

$$resp = ar_1^2 + br_2^2 \tag{4.6}$$

However, this model has the disadvantage in that it can only generate symmetric curves around the bases, whereas the sparse coding and Fan equation can

produce asymmetric curves. We believe that this asymmetry in the hyperselectivity is essential to efficiently represent the asymmetric distribution of natural scene data in the image state space.

The curvature produced by the cascaded linear-nonlinear models has interesting features. First of all, this model can produce both exo-origin curvature (hyperselectivity) and endo-origin curvature (invariance/tolerance). $resp = r_1 - r_2$ is an example of exo-origin curvature (shown in Figure 4.6j) and $resp = r_1 + r_2$ is an example of endo-origin curvature (not shown here). The other interesting difference is that curvature does not flatten out away from the vector representing the neuron.

Here we are not making a positive argument for a particular model. We just want to emphasize the curvature produced by each of these models and note the important differences and similarities. It has been demonstrated that the sparse coding network produces a variety of nonlinearities observed in V1 (Zhu and Rozell, 2013). The response geometry of the sparse coding network provides a deeper insight into these seemingly different nonlinearities. The fan equation produces curvature very similar to the curvature of the sparse coding network. With the fan equation, there is a deterministic relationship between the magnitude of curvature and the angle between the neighboring neurons. However, the fan equation model only works in 2D, although we are working towards expanding the fan equation to higher dimensions. The gain-control model also has been used to describe many nonlinear behaviors of a neuron (Tolhurst and Heeger, 1997; Mély and Serre, 2017). The cascaded linear-nonlinear models have the advantage that they can produce both selective and invariant curvature. Indeed, future work will be required to determine which model accurately deter-

mines the physiology.

In the next chapter, we will explore more about the sparse coding network in two-dimensional toy examples and higher-dimensional natural image state space. We quantify the curvature and how it depends on other free parameters of the network. Also, we will investigate the implications of curvature on hyperselectivity.

CHAPTER 5

THE CURVATURE OF THE SPARSE CODING NETWORK

In the previous chapter, we saw that the curvature in the iso-response contours provides a better description of a wide family of nonlinearities observed in V1. We saw that the sparse coding network curves the iso-response contours to produce these nonlinearities. In this chapter, we will explore in detail the inner workings of the sparse coding network and how different parameters of the network affect curvature. Later, we will also see how the hyperselectivity produced by the network breaks the Gabor-Heisenberg limit.

Since the discovery of the receptive fields of V1 neurons (Hubel and Wiesel, 1959, 1962, 1968), there has been tremendous effort put into understanding the physiological and computational properties of the neurons in the visual system. Barlow (1972) argued that the natural environment which has guided the evolution of the biological visual system is highly redundant and hence the neurons evolved to get rid of this redundancy. He argued that sensory neurons have organized themselves such that the few active neurons can reliably represent a stimulus completely. This idea that the sensory systems have evolved to efficiently represent the statistics of the natural environment is referred as the efficient coding hypothesis. Just as Gibson (1950) stressed that it is important to first understand the nature of environment before understanding the nature of the visual system that represents it, Barlow (Barlow, 1953, 1961; Barlow et al., 1967; Barlow, 1972, 1979), too emphasized the necessity of the study of redundancy in the natural environment to understand visual processing.

Following the efficient coding hypothesis of Barlows, Field (1987, 1994) analyzed the statistical redundancy in the natural scene images. He found that

two seemingly different natural scenes are very similar in statistical properties, whereas two random dot patterns which appear to be similar are extremely different from natural scene images. He quantified this redundancy in natural images using Fourier analysis. He found that the Fourier amplitude spectra of natural images are interestingly different from the flat Fourier amplitude spectrum of white noise images. He observed that natural images have greatest amplitudes at low frequency and the amplitude decreases as a function of increasing frequency. He further computed that the fall off in the amplitude is roughly proportional to the inverse of the frequency ($1/f$). This slope of the amplitude spectrum implies that there is a lot of redundancy in the neighboring pixels.

Following the argument about redundancy reduction by sensory systems, Atick and colleagues (Atick and Redlich, 1990, 1992; Atick, 1992; Dong and Atick, 1995; Dan et al., 1996) proposed a number of decorrelating (whitening) strategies employed by the retinal ganglion cells. However, this approach only removes the linear pairwise correlation (two-point correlation). Natural scene images consist of features such as edges, contours, curved edges, fractals which have higher-order correlations (e.g., three-point correlations) (Field et al., 1993; Olshausen and Field, 1997). It is important for any coding strategy to get rid of this redundancy as well.

Field (1994), found a linear coding strategy that takes into consideration these higher-order correlations. Field argued that Gabor-like filters found in V1 are efficient for representing natural scene images as they produce a sparse and distributed code, i.e., a very few numbers of active units (as argued by Barlow for sensory neurons). Field found that the distribution of the Gabor filter

responses to natural scenes was highly kurtotic. This means that for any given natural scene stimulus, only a few neurons respond, and on average the probability of any neuron responding is roughly the same. This implies that oriented Gabor filters can capture the higher-order correlations within natural images and hence require a fewer number of active neurons than any other linear coding strategy (e.g., PCA, Fourier).

To build an efficient representation system, one needs to identify the structures in natural scene images which have higher-order correlations (e.g., edges). However, this is analytically an intractable problem. To find structures of higher order correlations is not as easy as finding two-point correlation using Fourier analysis. Olshausen and Field (1996) developed a neural network could learn an efficient way to represent higher-order correlations. Field (1994), demonstrated that Gabor-like filters could capture some of these higher-order dependencies and thus produce a sparse and distributed code. Olshausen and Field (1996), thought of optimizing a neural network to learn a sparse and distributed code such that only a few responding neurons would be enough to represent any natural scene stimulus and every neuron would have an equal probability of responding overall. Olshausen and Field called this representation the sparse coding network and found that it learns Gabor-like basis functions similar to the receptive fields measured in V1. Later Olshausen (2013), found that ten times overcomplete sparse coding networks (i.e. ten times more encoding units than the input dimensionality; for example, a network learned with 64-pixel images as input will have 640 encoding neurons producing 640 outputs) learn basis functions which have structures other than the Gabor functions (e.g., Blobs, curved edges).

5.1 The sparse coding network

Generally, the sparse coding network attempts to reconstruct the input images. Given the feedforward weights (or basis functions) of M neurons, the j^{th} neuron ϕ_j is a vector of N dimensions which is the dimensionality of input image patches. All the ϕ_j's are set to have unit length. One can reconstruct back the input image patch from the outputs (a_j) of all the neurons.

$$\hat{I} = \sum_{j=1}^{M} \phi_j a_j \tag{5.1}$$

where $\hat{I}$ is the reconstruction of input image I. The only constraint on the network is that the outputs should be sparse (i.e., only a few of the neurons should be active (non-zero) for a given input image). The network is optimized by minimizing the following energy function (see Equations 5.2 and 5.3). The energy function tries to preserve information by computing the mean reconstruction error ($\frac{1}{2}\left|I - \sum_{j=1}^{M} \phi_j a_j\right|^2$) and sparseness of the outputs using cost function $S(a_j)$.

$$E = [preserveinformation] + \lambda \times [sparsenessofa_j] \tag{5.2}$$

$$E = \frac{1}{2}\left|I - \sum_{j=1}^{M} \phi_j a_j\right|^2 + \lambda \sum_{j=1}^{M} S(a_j) \tag{5.3}$$

The parameter λ controls the tradeoff between reconstruction and sparsity. A higher lambda value increases the sparsity in the network responses. $S(a_j)$ is the cost function which is a penalty function. It imposes penalty on the network whenever the network is not sparse, i.e. the cost on the network is more when there are many neurons active compared to the cost when only a few neurons are active. The cost function could be implemented using a simple L1 norm function. The popular choices of the cost functions are $abs(x)$, $\log(1 + x^2)$, and

$-\exp(-x^2)$. We will discuss more about the effect of the cost function in the next chapter.

The optimization of the network begins with a random set of feedforward weights and activities for each neuron. First, given the feedforward weights, the activities of the neurons are learned such that it minimizes the above energy function using the gradient descent. The activity of i^{th} neuron is learned using gradient with respect to a_i. The gradient is computed as follows:

$$
\begin{aligned}
-\frac{\partial E}{\partial a_i} &= -[I - \sum_{j=1}^{M} \phi_j a_j](-\phi_i) - \lambda S'(a_i), && (5.4)\\
&= I\phi_i^T - \sum_{j=1(j\neq i)}^{M} a_j[\phi_j^T \phi_i] - a_i \phi_i^T \phi_i - \lambda S'(a_i), && (5.5)\\
&= I\phi_i^T - \sum_{j=1(j\neq i)}^{M} a_j[\phi_j^T \phi_i] - a_i - \lambda S'(a_i), && (5.6)\\
&= b_i - \sum_{j=1(j\neq i)}^{M} G_{ij} a_j - f_\lambda(a_i) && (5.7)
\end{aligned}
$$

where $b_i = I\phi_i^T$, $G_{ij} = \phi_j^T \phi_i$ and $f_\lambda(a_i) = a_i + \lambda S'(a_i)$. If $S(a_i) = log(1 + a_i^2)$ then $S'(a_i) = \frac{2a_i}{(a_i^2+1)}$.

From this gradient step, we can see that the response of any neuron depends on the response of other neurons in the network (see Figure 5.1). The linear response of a neuron (b_i) gets inhibited by the weighted sum of the responses from other neurons ($\sum_{j=1(j\neq i)}^{M} G_{ij} a_j$) and the cost of sparseness $f_\lambda(a_i)$. The weighted sum depends on the dot product between each pair of the basis functions or the vectors representing the neurons (G_{ij}). If a pair of basis functions is orthogonal, then those two neurons will not inhibit each other as G_{ij} will be zero. Also, the inhibition would be stronger for neurons that have larger G_{ij} values. This implies that a pair of basis functions (neurons) which have an angle less than 90 degrees between the vectors, will take part in inhibition, and the strength of

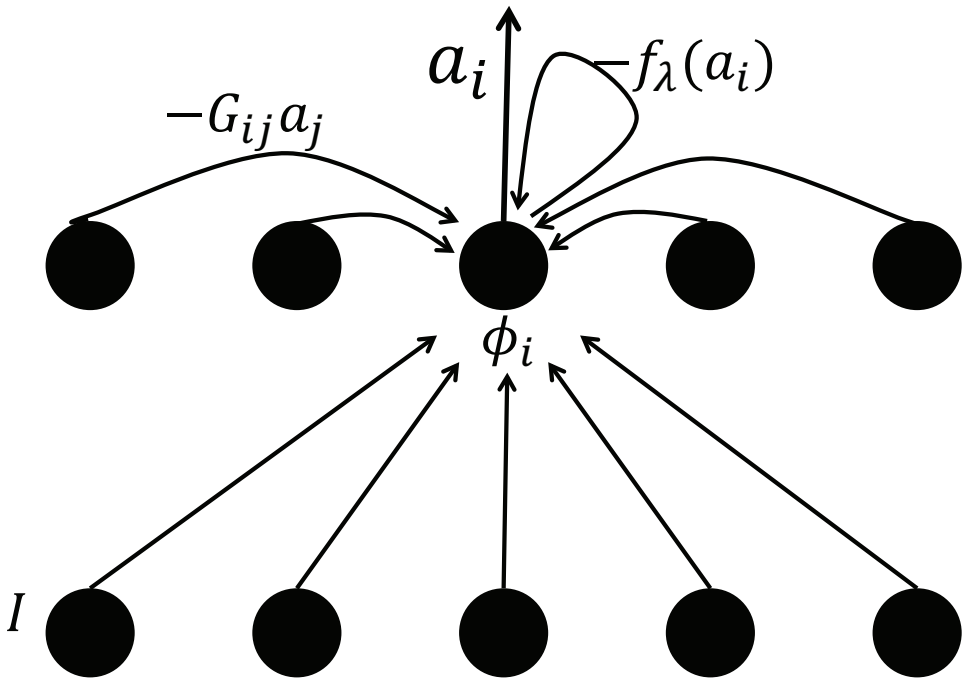


Figure 5.1: The figure shows the network diagram adapted from Olshausen and Field (1997). The output of neuron ϕ_i is inhibited by other neurons ($G_{ij}a_j$) and cost of sparseness $f_\lambda(a_i)$.

inhibition is directly proportional to the angle between the vectors representing the neurons.

Once the activities of neurons are optimized, the next step is to learn the optimal feed-forward weights of all the neurons in the network given their learned activities in the previous step. The weights for the i^{th} neuron ϕ_i are learned using the gradient of the energy function with respect to ϕ_i. The gradient is computed as follows

$$-\frac{\partial E}{\partial \phi_i} = [I - \sum_{j=1}^{M} \phi_j a_j]a_i \tag{5.8}$$

These two steps of learning activities and feed-forwards weights are repeated one after another until there is no significant decrease in the energy function. Figure 5.2 shows the learned basis functions from a 1.3× sparse cod-

ing network, where the number of output neurons are equal to 1.3 times the number of input pixels (i.e. $M = 1.3 \times N$). The basis functions learned are oriented Gabor functions similar to the receptive fields found in V1. Figure 5.3 shows the basis functions learned from a 13 times overcomplete network (i.e., $M = 13 \times N$). Notice that a highly overcomplete network finds basis functions other than Gabor-like functions. There are many other forms learned such as spots, curves, and plaids. The ability of the sparse coding network to learn overcomplete codes is one of the advantages over the early versions of Independent Component Analysis (ICA), which also learns the linear sparse and distributed codes. However, there are now overcomplete ICA algorithms (e.g., Lewicki and Sejnowski (2000)).

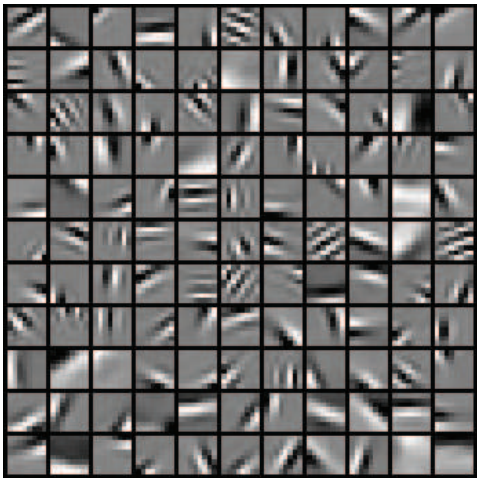


Figure 5.2: The basis functions learned from a 1.3× overcomplete sparse coding network.

So far, all the efficient coding models of the visual system focused on the basis functions learned from the statistics of natural scenes. The 2D maps of the basis functions (Figure 5.3), which show the features they are selective to, are important, but that does not demonstrate the nonlinear response behavior. Golden et al. (2016) for the first time observed that some of these efficient coding

techniques (e.g., the sparse coding model) produce curvature in the response geometry of the neurons. And, interestingly the family of curvature the sparse coding model produces is the same family which describes a wide family of nonlinear behavior observed in V1. It has been demonstrated that the overcomplete sparse coding network learns a nonlinear representation that gives rise to the well-known nonlinearities such as end-stopping, cross-orientation inhibition and other non-classical receptive field effects (Zhu and Rozell, 2013; Lee et al., 2006). Here, I will show first the curvature produced in two-dimensional toy data to allow an intuitive understanding and then the curvature in high-dimensional (64-dimensional) natural scene data using the sparse coding network model.

5.2 Exo-origin curvature in the two-dimensional sparse coding network

Here we will analyze the exo-origin curvature in iso-response contours produced by the two-dimensional sparse coding network. For this we created a 2D sparse dataset with three causes of data, which means that most of the data lie along the three directions in the image state space. Figure 5.4a shows a scatter plot of the dataset in two-dimensional image state space. Most of the data lie sparsely along the three directions. We trained a sparse coding network to learn these data using three neurons. Figure 5.4b shows the learned directions of the three neurons (represented as black vectors) in the network. We can see that the network learns to arrange the neurons in the direction of the causes of the data (i.e., the directions along which the data lies in the image state space). Fig-

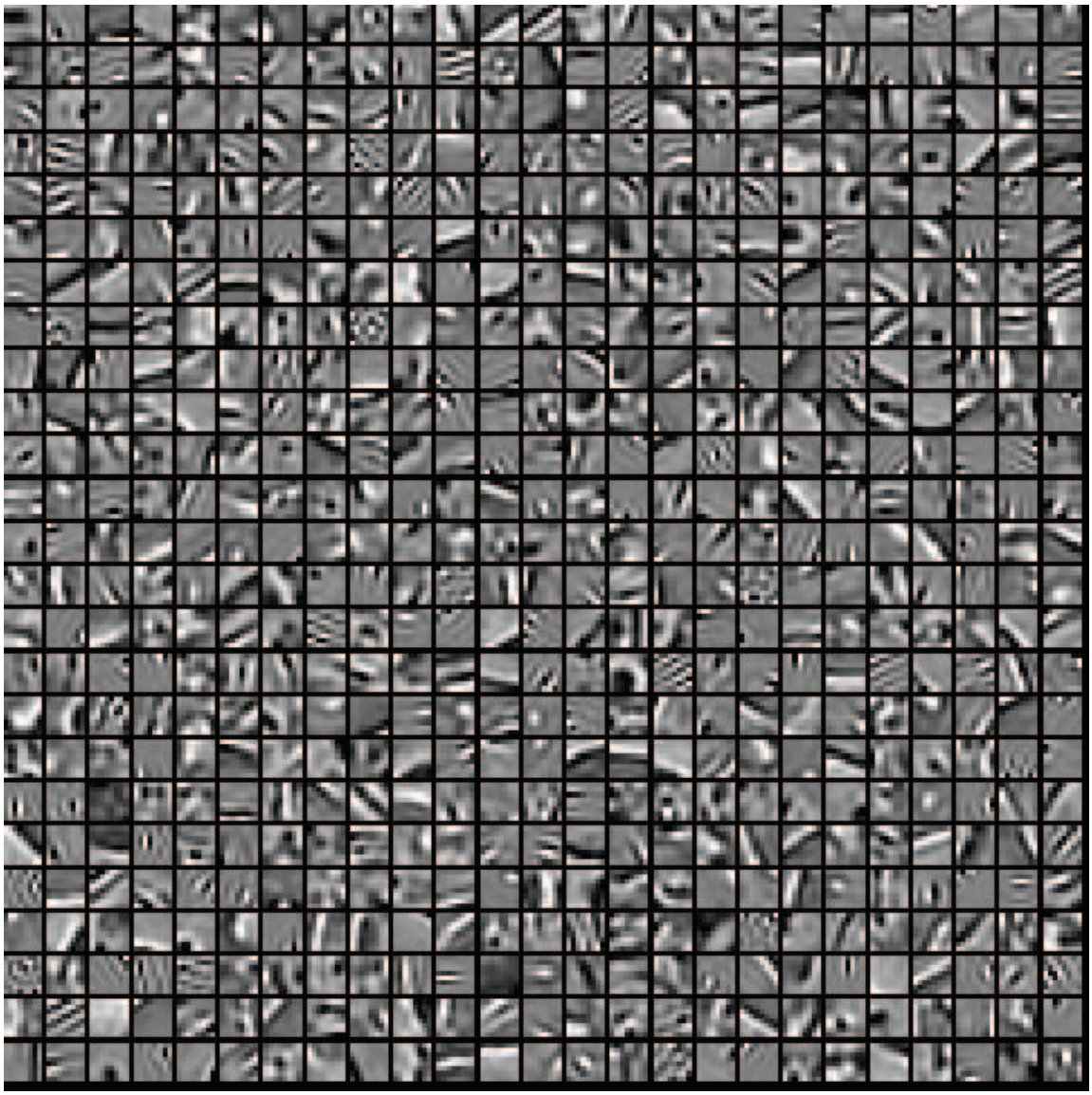


Figure 5.3: The basis functions learned from a 13× overcomplete sparse coding network.

ure 5.4b and c show the results from two different networks which use different values of λ. As we have discussed previously, the λ parameter determines the sparseness in the network representations. A higher λ value produces a sparser representation than a smaller λ value. To see the effect of λ on the network response, we plotted the iso-response contours of each neuron (iso-response con-

tours are plotted using three colors representing the contours of each neuron) by evaluating the response of each neuron at a grid of data points uniformly distributed over the 2D image state space. For the results of Figure 5.4b, the λ was set to a small value of 0.001. With a small λ value, we do not see much curvature in the iso-response contours. These neurons are equivalent to any linear neuron. However, with a large λ value of 0.25 (Figure 5.4c) we clearly see curvature in the iso-response contours of all three neurons.

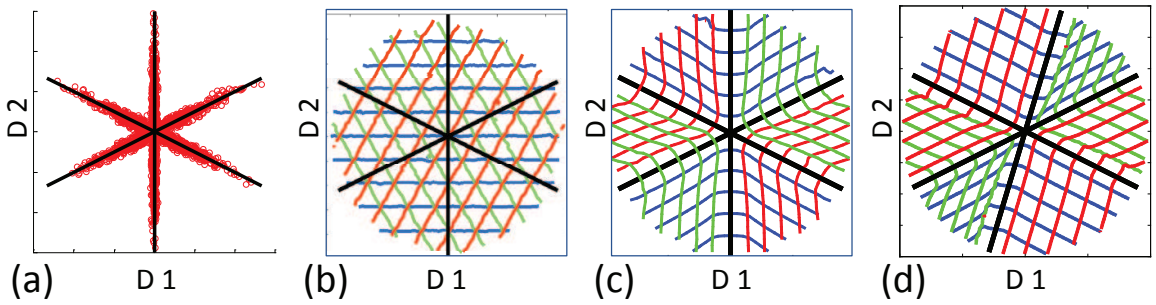


Figure 5.4: The iso-response contours from an overcomplete sparse coding network in 2-dimensional image state space. (a)Scatter plot of 2D sparse data with three sparse causes represented by the three axes. (b) and (c) Results of the sparse coding network with three basis vectors (1.5× overcomplete). The plots show the iso-response contours for each of the three neurons. (b) shows the result when $\lambda = 0.01$. (c) shows the result when $\lambda = 0.25$. With higher λ the network puts more emphasis on finding a solution that is sparse. The network's representation is a result of a recurrent nonlinear computation. As one can see, the iso-response contours have exo-origin curvature. This results in a representation where no more than two neurons are active for any given stimulus. Iso-response contours of each neuron are shown with different colors. (d) shows the result when the causes are not symmetrically distributed. As one can see the curvature that is learned is asymmetric. However each region of the space is represented by no more than two neurons.

We argue that solution shown in Figure 5.4c is much more efficient than the solution shown in Figure 5.4b. We can see that in Figure 5.4b, for any data point

in the image state space all the three neurons will produce a response. An efficient solution in a two-dimensional image state space should only have two neurons responding to any given data point. However, because of the overcomplete nature of the network (three neurons in 2D image state space), there is an inherent redundancy in the response of the network which results in a non-sparse solution. To get rid of this redundancy one needs to increase the sparseness of the solution by increasing the λ value. By increasing the sparseness, the network learns to produce the non-linear responses which curve the iso-response contours (as shown in Figure 5.4c). Due to of this curvature, any data point in the image state gets represented by only two neurons. That is for any stimulus no more than two neurons are active, which is an efficient solution. We call this solution "critically sampled overcomplete". In general "critically sampled overcomplete" means, when n-dimensional data are represented by k neurons (vectors), where k is larger than n ($k > n$), then only n neurons respond to any given stimulus.

5.3 Exo-origin curvature in high-dimensional image state space

In this section, we will visualize the exo-origin curvature produced by the sparse coding network in high-dimensional image state space. We trained a two times overcomplete sparse coding network on high-dimensional (8×8 - pixel) natural scene image patches. The network was initialized with 128 neurons, twice the input dimensionality. Figure 5.2 shows the learned 2D spatial map of the basis functions (neurons). This is the typical result that you get out of a sparse coding network trained on high-dimensional natural scene images. Here we will visualize the response behavior of these neurons in high dimen-

sional space. However, we cannot visualize the entire 64-dimensional geometry of a neuron's response. Instead, we will visualize two-dimensional subspaces between the neurons. Two visualize the curvature in iso-response contours, we probed the network with 2D subspaces defined by every possible pair of neurons. Each point in these subspaces corresponds to a 64-dimensional image patch. To map out the responses in each subspace, the sparse coding network is probed with a uniform grid of $64D$ data points in the subspace. Figure 5.5a and b show such two pairs and the iso-response contours in corresponding 2D subspaces. In Figure 5.5a we see a pair where the two basis functions have no overlap and hence are orthogonal in image state space (shown as vectors in red). As we have seen previously, the orthogonal neurons do not inhibit each other (see Equation 5.4) and hence produce either no curvature in its iso-response contours ,or only a small amount. Figure 5.5b shows a pair where there is significant overlap between the 2D spatial map of the learned basis functions. This overlap results in an angle of 60 degrees in the image state space (shown as red vectors in Figure 5.5b). This non-orthogonal overlap produces nonlinear inhibition and causes exo-origin curvature in the iso-response contours.

5.3.1 Measuring the curvature using parabolic fits

We quantify the curvature in iso-response contours by fitting a parabola ($y = ax^2 + b$) to the iso-response contour of magnitude 0.1. In higher dimensions, measuring the curvature in a neuron's response geometry is not a straightforward computation. There are an infinite number of directions in which one could probe the network to quantify the curvature. However, we believe that in most of the directions the response surface of a neuron is flat, as most of the

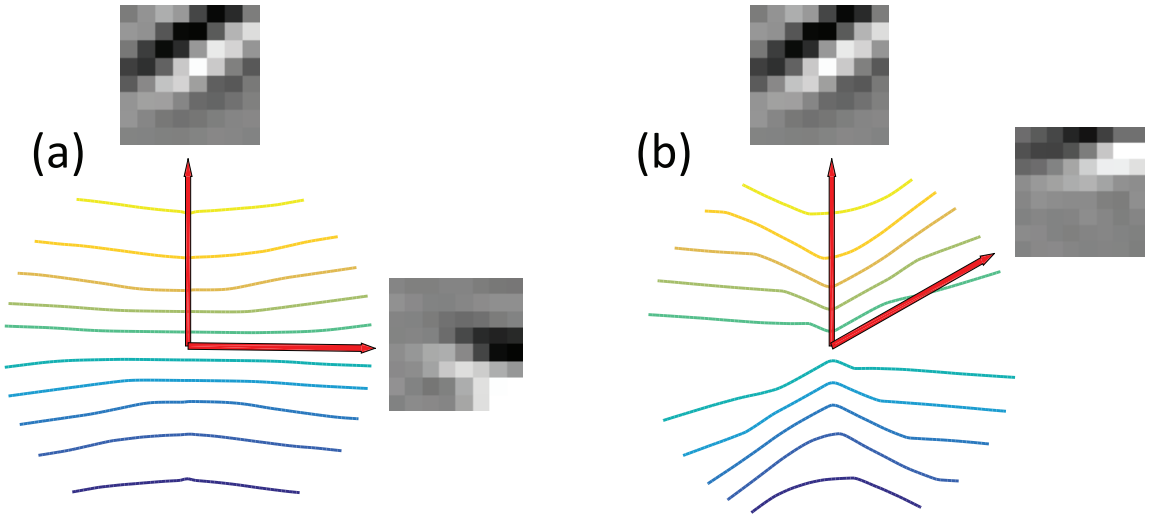


Figure 5.5: The iso-response contours in high-dimensional sparse coding network. The figure shows 2D subspaces between two neurons represented by the vectors. (a) shows an example of two basis functions (neurons) from the learned basis set that are orthogonal. The iso-response contour is shown for one of the neurons (represented by the vertical vector). Since the neighboring vector is orthogonal, we do not see any curvature in the resulting iso-response contour. (b) shows an example of two basis functions (neurons) from the learned basis set that have 60 degrees of angle between the vectors in the image state space. The iso-response contour is shown for one of the neurons (represented by the vertical vector). Since the neighboring vector is less than 90 degrees away, we see curvature in the resulting iso-response contour because of the inhibition from the neighboring vector.

directions do not have any neighboring neurons to influence the response. It is more interesting to measure the curvature in the directions where there are neighboring neurons. Hence, here we will focus only in the directions where there are neurons. For the computation of curvature, if a network has n neurons then we will consider every pair ($n(n-1)/2$ pairs) of neurons and probe the 2D subspace defined by each pair. In each 2D subspace, we select the single iso-response contour that represents the response magnitude of 0.1. The direc-

tion of the vector representing the neuron is aligned with the y-axis, and the iso-contour is fit with the equation of the parabola ($y = ax^2 + b$). Since the iso-response contours can be asymmetric along the direction of the vector, we fit the parabola only to the part of the iso-response contour between the two vectors (i.e., the vector representing the neuron under analysis and the vector representing the neighboring neuron). The magnitude of the curvature is defined as the value of the parabolic fit parameter a.

Figure 5.6 shows the magnitude of curvature as a function of angle and overcompleteness in the network. The four subplots of Figure 5.6 show the curvature in the sparse coding network of four different degrees of overcompleteness (1×, 1.3×, 2.6× and 5.2×). Each plot shows the magnitude of curvature (parabolic fit parameter a) on the y-axis as a function of the angle between the two neurons on the x-axis. All the four plots show that the curvature increases as the angle between the neighboring neurons decreases. The red line is the linear fit to the data points. We can see that the linear fits (red lines) get steeper (i.e., slope increases) as the overcompleteness increases. With a higher degree of overcompleteness, there are more neuron pairs that have angle less than 90 degrees; this produces more non-linear inhibition which results in more curvature in the iso-response contours. The solid black curves in each plot represent the predicted curvature from the fan equation model (see Chapter 4) in the 2D subspace. We can see that the curvature prediction from the fan equation is same irrespective of the overcompleteness because the curvature in the fan equation depends only on the angle between the neighboring neurons. Also, we can notice that the curvature from the sparse coding network is less than the predicted curvature from the fan equation. We believe that for some reason the sparse coding network is not achieving the most efficient solutions in higher dimensions.

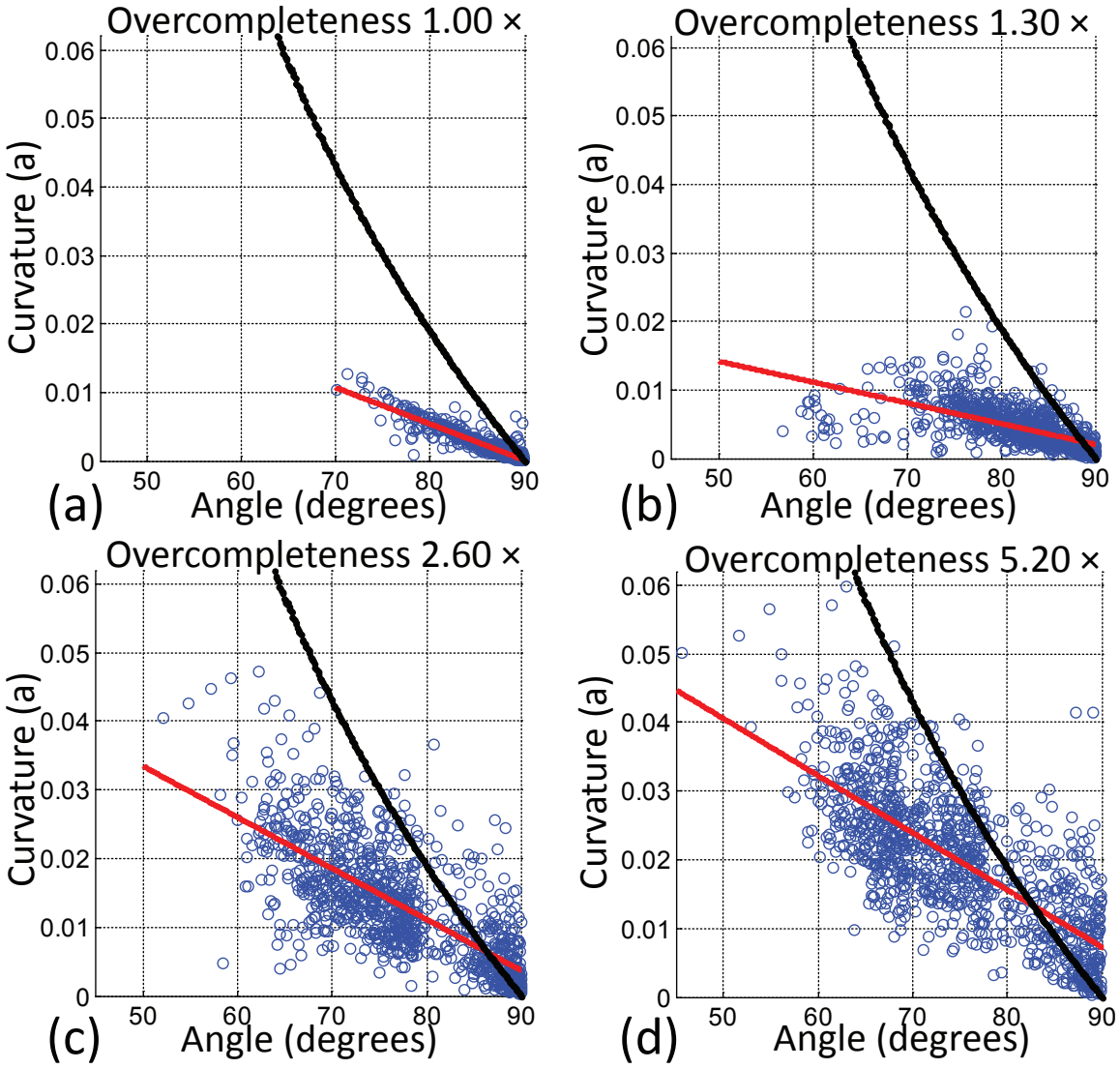

Overcompleteness 1.00 ×
Overcompleteness 1.30 ×
Overcompleteness 2.60 ×
Overcompleteness 5.20 ×
Curvature (a)
Angle (degrees)
0
0.01
0.02
0.03
0.04
0.05
0.06
50
60
70
80
90
(a)
(b)
(c)
(d)

Figure 5.6: The four figures show the curvature of the iso-response contours for the sparse coding network when trained on natural scenes. Curvature was measured in the two-dimensional subregions defined as the region between each pair of learned neurons (vectors). Results are shown for four degrees of overcompleteness using a measure of parabolic parameter a (see text). Note that the curvature is at a minimum for vectors that are orthogonal (90 degrees). For angles less than 90 degrees the curvature increases (higher exo-origin curvature) with decreasing angle. As the representation becomes more overcomplete we find more neurons with a high degree of curvature. The red line shows a linear fit to the data, and the increasing slope of the line with overcompleteness of the network indicates that curvature generally increases with overcompleteness. The black lines in each of the figures shows the predicted curvature of the iso-response contours generated using the fan equation (Equation 4.4). As one can see the curvature with the sparse coding network is less than that predicted by the fan equation. This figure is re-plotted from Golden et al. (2016).

5.3.2 Curvature vs. overcompleteness

As we noted previously, models of efficient coding focused mainly on the 2D spatial maps of basis functions (receptive fields that resembled Gabor-functions). However, this ignores the response behavior of these neurons. We have seen the receptive fields learned from 1.3 times overcomplete and 13 times overcomplete sparse coding network. These learned receptive fields appear to be similar to the standard Gabor-like functions (except some receptive fields in 13 time overcomplete network, e.g., spots, curves, and plaids). Interesting differences with the highly overcomplete networks have been observed, but we believe that more interesting differences are in the response geometry of these overcomplete networks. Two neurons with similar receptive fields from two dif-

ferently overcomplete networks can have significant differences in the curvature of their iso-response contours. Figure 5.7a and b, show the effect of overcompleteness on the curvature. When the network is one-times overcomplete, the bases are orthogonal, producing little or no curvature (see Figure 5.7a). However, when the network is 1.5 times overcomplete, there are more neurons than the dimensionality. This produces nonlinear inhibition between the neurons, causing the iso-response contour to warp. One should note that in 2D image-state space, increasing the overcompleteness has a larger effect on the angle between the neurons. Increasing the number of neurons from 2 to 3, decreases the angle between the vector from 90 degrees to 60 degrees (assuming the vectors are uniformly distributed in the state space). However, in higher dimensions, increasing the number of neurons has a relatively smaller effect on the average angle between the vectors. This implies that on average there is a small effect on the curvature as well. To understand the effect of overcompleteness on the curvature, we will only consider the pairs of neurons with the smallest angles.

Figure 5.7c shows the average smallest angle between pairs of learned neurons and the curvature as the function of overcompleteness in the sparse coding network. We used 8×8 natural scene image patches to train sparse coding networks with overcompleteness ranging from 1.3 times to 13 times. In each network, we computed the angle between each pair of the learned bases and the curvature of the iso-response contour in the 2D subspace defined by the pair of basis vectors. The figure shows the average angle and the curvature of the five closest basis vectors to each basis. This plot shows the average curvature in the most curved region of the image-state space. These results indicate that with an increase in the degree of overcompleteness, the angle between neighbors decreases and the curvature(hyperselectivity) increases.

In the next chapter we will see how the curvature produced by the sparse coding network makes a neuron hyperselective and breaks the Gabor-Heisenberg limit.

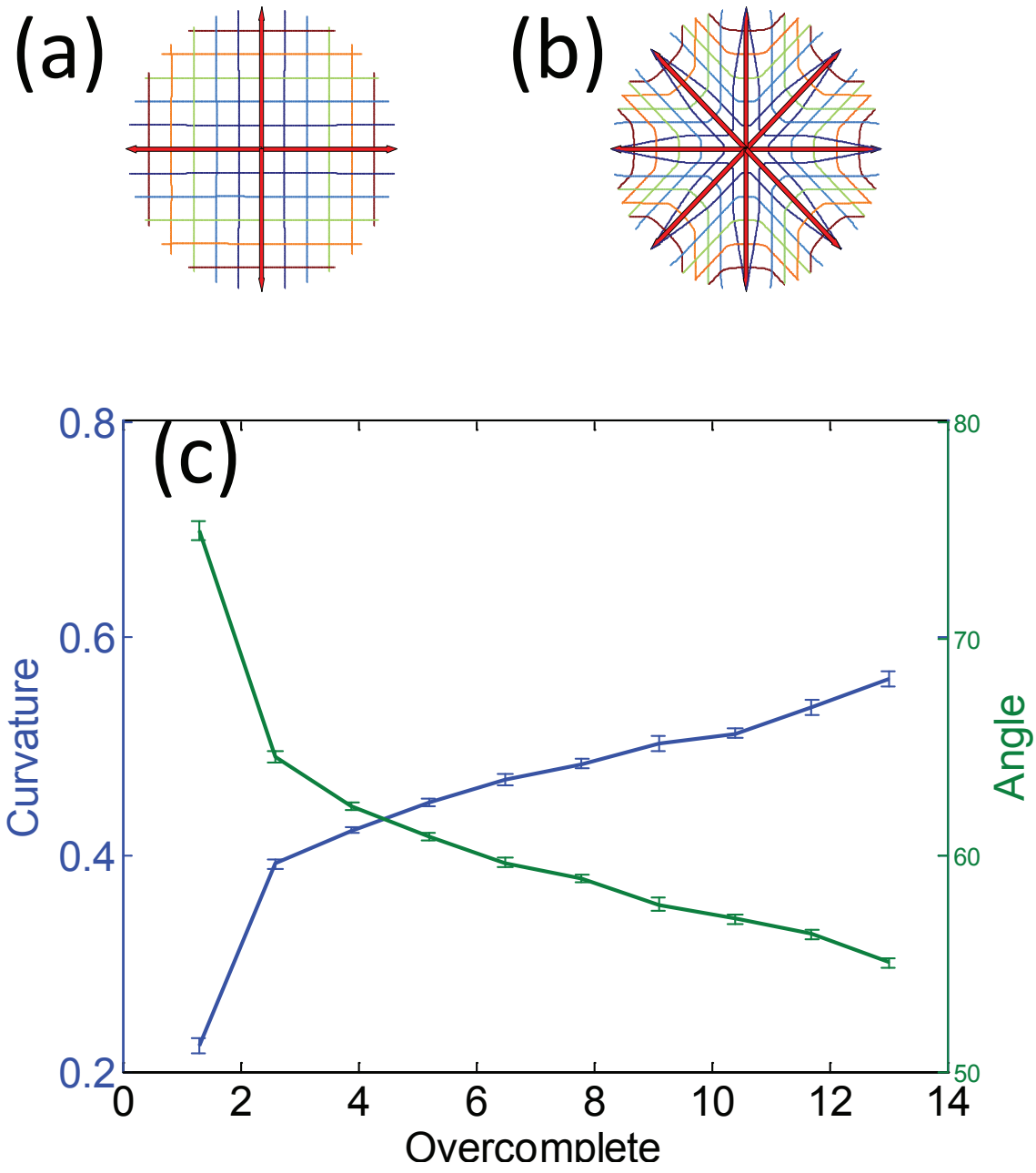


Figure 5.7: (a) shows an example of the iso-response contours in when the neighbors are orthogonal. In an overcomplete network there are more more neurons than dimensions(e.g., pixels). This forces the angles between many neurons to be less than 90 degrees. (b) shows the curvature in $2D$ space when there are four neurons representing that space. (c) shows the curvature changes as the sparse coding network become more overcomplete. For this figure, we trained a sparse coding network on 8×8 natural scene image patches. We varied the overcompleteness of the network from 1.3 times(e.g., Olshausen and Field (1996)) to 13 times. We then measured the curvature for the $2D$ subspace defined between any pair of neurons in the network. For all of these networks the majority of pairs will be orthogonal. We therefore measured the curvature for only the five neurons with the most overlap for each neuron in the network(i.e., the five neurons with smallest angle in the image space). See text for details. (c) plots the average curvature as a function of overcompleteness. The figure also shows average smallest angle of these five closest bases for each bases as function of overcompleteness. As one can see as the network becomes more overcomplete the curvature between neighbors increases(i.e., the network becomes more hyper-selective).

CHAPTER 6

HYPERSELECTIVITY

As noted in previous chapters, the physiologists and the modelers of the visual system were hystorically focused on determining the visual feature (receptive field) to which a neuron is selective. Since the early recordings of the visual neurons (e.g., Kuffler (1953); Barlow (1953); Hubel and Wiesel (1959)), researchers have probed neurons with a wide variety of stimuli to determine the receptive fields. The mapping out of the response characteristics of a neuron to spot or line stimuli as a function of position provides a description of the neuron. For linear neurons, this description is complete, and one can easily predict the response of the neuron to any given novels stimulus as a simple dot product between the receptive field ($\vec{rf}$) of the neuron and the stimulus (S). However, real neurons are highly nonlinear.

$$R(S) =< \vec{rf}.S > \tag{6.1}$$

Based on receptive fields measured in physiological experiments, a variety of models have been developed which predict the response of a neuron to a simple stimulus accurately but fail to predict the response to naturalistic stimuli (e.g., Prenger et al. (2004); Olshausen and Field (2005); Murray (2011); David and Gallant (2005); Mante et al. (2005)). Although, it is well known that the neurons are highly nonlinear, the receptive fields of neurons are commonly considered as a description of their selectivity. Multiple attempts have been made to model these nonlinearities, which can account for responses to complex stimuli such as natural scenes.

Here in this chapter, I will demonstrate that the receptive field only is not the complete description of a neurons selectivity. Also, I will show how curva-

ture in the iso-response contours affects the selectivity of neurons causing them to become hyperselective (Vilankar and Field, 2017). We define two forms of selectivity. The first type is the "classic selectivity". In this kind of selectivity, a neuron produces the maximum response to an optimal stimulus (S_{max}). If a neuron is linear, then the optimal stimulus (receptive field) is determined by probing the neuron with an orthogonal basis set (e.g., spots or gratings). The second form of selectivity we described as "hyperselectivity". This is the measure of how narrowly tuned a neuron is around its optimal stimulus. A neuron is hyperselective if the neuron's response to a hybrid stimulus $S_{max} + S_2$ (where S_2 is orthogonal to the optimal stimulus S_{max}) is less than the response to the optimal stimulus S_{max}. This hyperselectivity tuning around the optimal stimulus implies exo-origin curvature in the iso-response contours.

$$R(S_{max} + S_2) < R(S_{max}) \mid S_{max} \perp S_2 \tag{6.2}$$

I will differentiate between a neuron's optimal stimulus and the selectivity around the optimal stimulus (hyperselectivity). We will demonstrate that because of the curvature, the estimate of the receptive field is not the neuron's optimal stimulus. We will show that the spatial frequency bandwidth is much narrower than the predicted receptive fields. We will show a paradoxical phenomenon where a neuron can be narrowly tuned to a broadband stimulus. Finally, we will show that response curvature can cause a neuron to break the Gabor/Heisenberg limit (i.e. to have hyperselectivity in both the frequency and space below the optimal selectivity of Gabor functions).

6.1 Classical concept of selectivity

Historically, a number of theories have been proposed regarding the function of the neurons in the early visual system. They have ranged from basic edge detection (Marr and Hildreth, 1980) to suggestions that the visual system carries out something like a Fourier transform (e.g., see De Valois et al. (1978)). The optimal stimulus or the receptive field was determined by probing a neuron with an orthogonal basis set. The optimal stimulus was then simply computed as the weighted sum of the inputs.

$$S_{max} = \sum_{i=1}^{n} \psi_i * R(\psi_i) \tag{6.3}$$

where, ψ is any orthonormal basis set, $R(\psi_i)$ is the linear response to stimulus ψ_i.

Figure 6.1 shows the receptive field of a neuron (on the left) and the response profile as a function of spatial frequency using grating stimuli (on the right). For a linear neuron, the two plots in Figure 6.1 are simply the Fourier transform of each other. Since the response of this neuron is localized in the frequency domain classically, this neuron will be considered a narrowly tuned neuron. However, we argue that the selectivity of this neuron is not different from that of any other linear neuron. This neuron is simply a vector in high-dimensional state space and is equally selective to the stimuli around it as any other linear neuron. Figure 3.3 a) shows such a linear neuron in low-dimensional state space. The iso-response contours for the linear neuron are straight and perpendicular to the optimal stimulus (S_{max}) directions (shown as a black vector). For a linear neuron, any orthogonal stimulus $S2$ added to the optimal stimulus will produce no change in the response.

$$R(S_{max} + S_2) = R(S_{max}) \mid S_{max} \perp S_2 \tag{6.4}$$

This equation still holds true for neurons with planar nonlinearities because they have iso-response contours that are straight and perpendicular to the optimal stimulus.

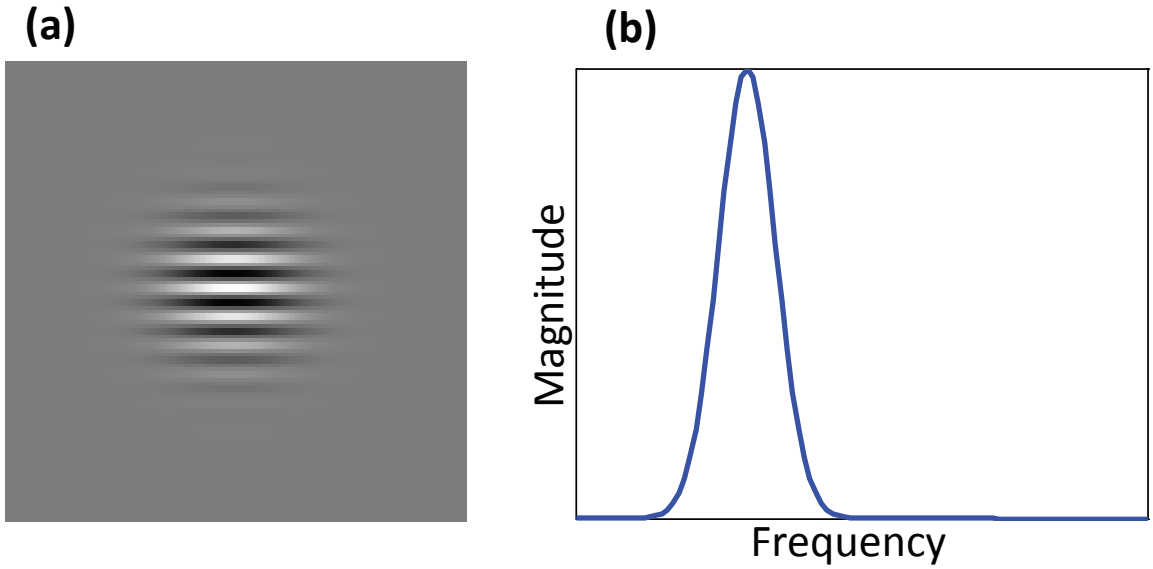


Figure 6.1: (a) shows a neuron's receptive field. Classically this neuron would be considered a "narrowly-tuned" neuron, because of its localized magnitude response in frequency domain (shown in (b)).

6.2 Hyperselectivity

Here we define a new concept of selectivity that we termed "hyperselectivity" (Golden et al., 2016; Vilankar and Field, 2017). Hyperselectivity is defined by the response selectivity region in a neurons response geometry in the image state space. This selectivity falls out of the curvature in the iso-response contours. Unlike the classical concept of the selectivity, hyperselectivity does not depend on the optimal stimulus (direction in the image-state space). The hyperselec-

tivity of a neuron depends on the nonlinear inhibition from the neighboring neurons, the overcompleteness of the network and the sparseness factor (as discussed in the previous chapter and Golden et al. (2016). Consider Figure 3.4a, which shows a neuron pointing in its optimal direction with exo-origin curvature in its iso-response contours. The exo-origin curvature makes this neuron hyperselective with greater selectivity than that of a linear neuron. This neuron responds only to a limited volume in the state space, and the magnitude of the hyperselectivity depends on the curvature of the iso-response contours. A neuron is hyperselective if the response to the optimal stimulus (S_{max}) plus an orthogonal stimulus (S_2) is less than the response to S_{max} (Equation 6.2).

In Chapter 3 we discussed four approaches/models that produce exo-origin curvature and thus generating hyperselectivity: sparse coding (e.g., Olshausen and Field (1996)), the fan equation (Golden et al., 2016), gain control with divisive normalization (e.g., Heeger (1992); Schwartz and Simoncelli (2001)), and a recent example of a linear non-linear model (Pagan et al., 2016). All these models have free parameters that can affect the hyperselectivity of neurons in the network.

6.3 The effect of curvature on orientation bandwidth tuning

Classically, a neuron's selectivity is determined from its response tuning in spatial frequency and orientation. However, describing a neuron in such a way could be misleading. While it is perfectly fine for a linear neuron, this is not an appropriate description for neurons with response curvature. The curvature can produce a paradoxical neuron we call "narrowly tuned to a broadband

stimulus". Figure 6.2a) shows a receptive field of a neuron, the receptive field resembles with a vertical wide Gabor (low spatial frequency) function. Traditionally, this neuron would be considered a broadband neuron because of its broad tuning in spatial frequency and orientation. If this neuron is a linear neuron, then it will have no curvature and will produce straight iso-response contours as shown in Figure 6.2d). The response of this linear neuron to gratings of various orientation will produce the orientation bandwidth profile as shown in red in Figure 6.2g) and h).

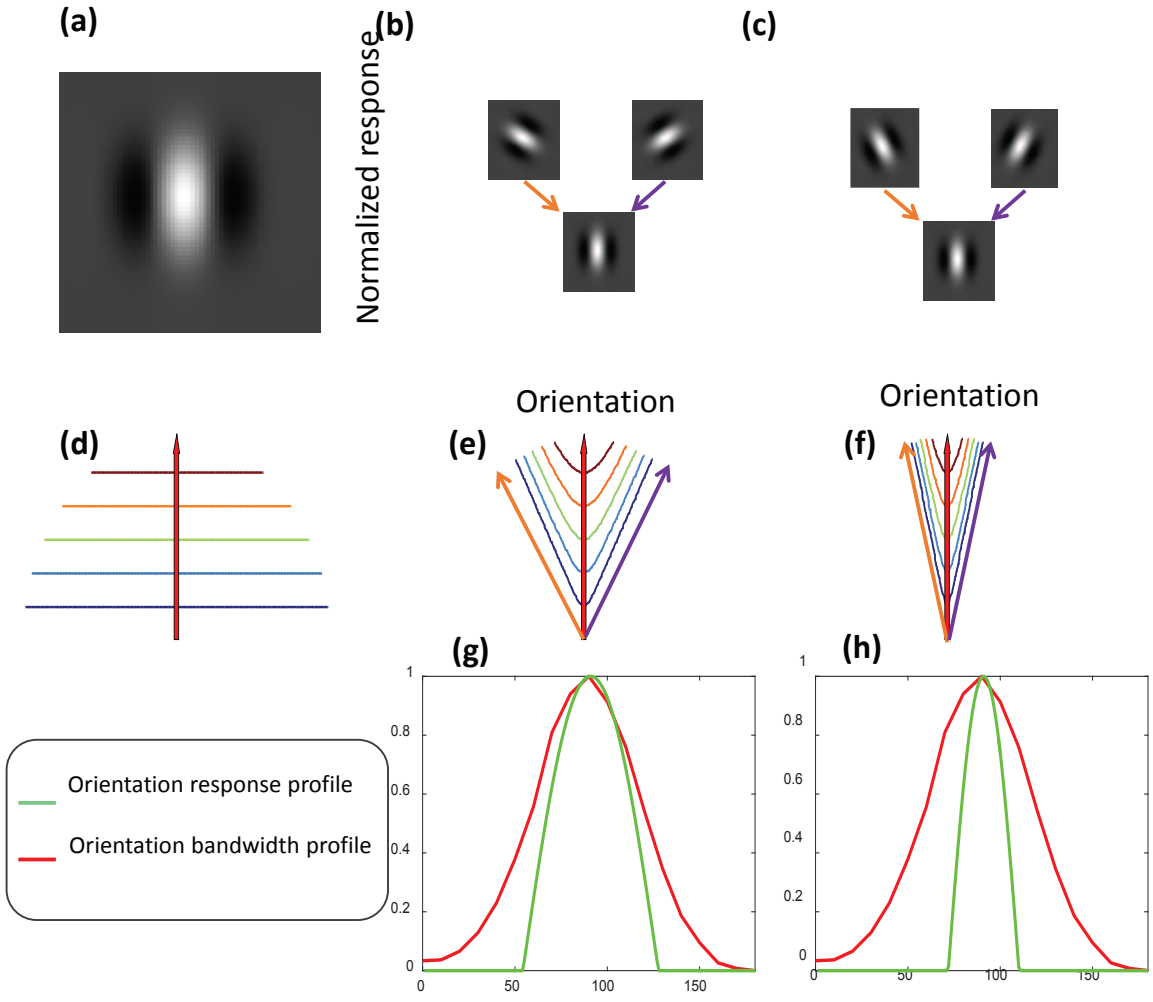


Figure 6.2: (previous page): The figure shows that the hyperselectivity can produce a paradoxical neuron that is narrowly tuned to a broadband stimulus. (a) shows the receptive field of a neuron that would classically be considered as broadband. The green curves in (g) and (h) show the effects of the nonlinear inhibition by two neighbors with similar orientations (the inhibitory neurons are shown in (b) and (c) respectively). In the scenario shown in (b), the vertical oriented neuron is inhibited by the neighboring neuron with orientations of 60 and 120 degrees. In the scenario shown in (c), the vertical oriented neuron is inhibited by the neighboring neurons with orientations 75 and 100 degrees. The response of the neuron was modeled by using the Fan equation. (e) and (f) shows the curvature that is produced with these neighbors. With this curvature, the optimal stimulus is unchanged. However, it responds less to nearby orientations. If the neuron is mapped with stimuli of different orientations then the neuron will appear to be narrow band. However, its preferred stimulus has not changed. The curvature produced by these neighboring neurons interaction allows the neuron to be highly selective to this broadband stimulus(i.e., its optimal(S_{max}) is still the broadband Gabor function shown in a)).

Now consider the scenario shown in Figure 6.2b), where the neuron shown in Figure 6.2a) is now flanked with two more similar neurons (same in spatial frequency) with slightly different orientation in their optimal stimuli. These neighboring neurons are oriented at 60 and 120 degrees (shown in Figure 6.2b). The orientation difference is of 30 degrees with the center vertical neuron. In the image state space, the flanking neurons (shown as orange and purple vectors in Figure 6.2e) would be closer to the neuron in the center (shown as a red vector). The presence of these neurons would produce non-linear inhibition causing curvature as shown in Figure 6.2e). This curvature would not change the optimal stimulus; the optimal stimulus would still be the wide Gabor shown in Figure 6.2a). However, the response profile for gratings of various orientations would change producing a narrow tuning in orientation as shown by the green

curve in Figure 6.2g. If the flanking neurons are oriented at 75 and 105 degrees (shown in Figure 6.2c), there would be more curvature in the iso-response contours (shown in Figure 6.2f), producing an even narrower orientation response profile (shown in Figure 6.2h). However, the optimal stimulus would be the same wide Gabor function. Thus, with curvature, it is possible to get a neuron narrowly tuned to a broadband stimulus.

6.4 Incorrect estimation of the optimal stimulus (receptive field)

Generally, in physiology experiments for the estimation of the receptive field (i.e. the optimal stimulus S_{max} that produces the maximum response) is performed by probing a neuron with an orthonormal basis set. However, given the nonlinearity (curvature), the estimation of the optimal stimulus could be misleading. I will demonstrate that the choice of the orthonormal basis set (e.g., spots and gratings) influences the estimation of the optimal stimulus. We will see that the estimated optimal stimulus depends on the choice of the orthonormal basis set, and that the receptive field obtained from two different basis sets are different from each other. Further, we will quantify the error between the estimated and the true receptive field in terms of angular distance the image state space.

The discrepancy in the estimation of the receptive field has been noted in the physiology before. The tuning of a neuron's response, when measured with pixels or lines, is found to be different from the tuning when measured with gratings (e.g., Tadmor and Tolhurst (1989); Tolhurst and Heeger (1997)).

These differences have been attributed to non-linearities such as threshold non-linearities (e.g. see Tadmor and Tolhurst (1989); Tolhurst and Heeger (1997)), contrast gain control (e.g., Tolhurst and Heeger (1997)) , frequency suppression (De Valois et al., 1985) or some form of surround suppression (e.g., Nestares and Heeger (1997)). Here we show that we can observe this effect in the two-times overcomplete sparse coding network (this effect was briefly described earlier by Olshausen and Field (1997)). We will demonstrate this extensively in the complete network and estimate on average how far the estimation of the receptive field is from the true optimal stimulus.

Figure 6.3a shows the basis functions (i.e., the feed-forward weights) learned using two-time overcomplete sparse coding network. These basis functions are the true optimal stimuli of the network. Ideally, if the network is linear, then irrespective of the probing orthonormal basis set, one should find the estimated receptive field to be same as that shown in Figure 6.3a. However, because of the overcomplete nature of the network, neurons become hyperselective. This causes the estimate of the receptive field to be dependent on the choice of the orthonormal basis used to measure the neuron's response. Figure 6.3b and c show the estimated receptive fields using two different sets of orthonormal basis set. Figure 6.3b shows the receptive fields estimated using spot (pixel) basis set. Each neuron's receptive field was estimated by measuring the response as a function of pixel position. Figure 6.3c shows the receptive fields estimated using sinusoidal grating basis set. Each neuron's receptive field was estimated by computing the inverse Fourier transform of its response spectrum. We can observe that the receptive fields estimated using pixels are smaller than the receptive fields measured by gratings.

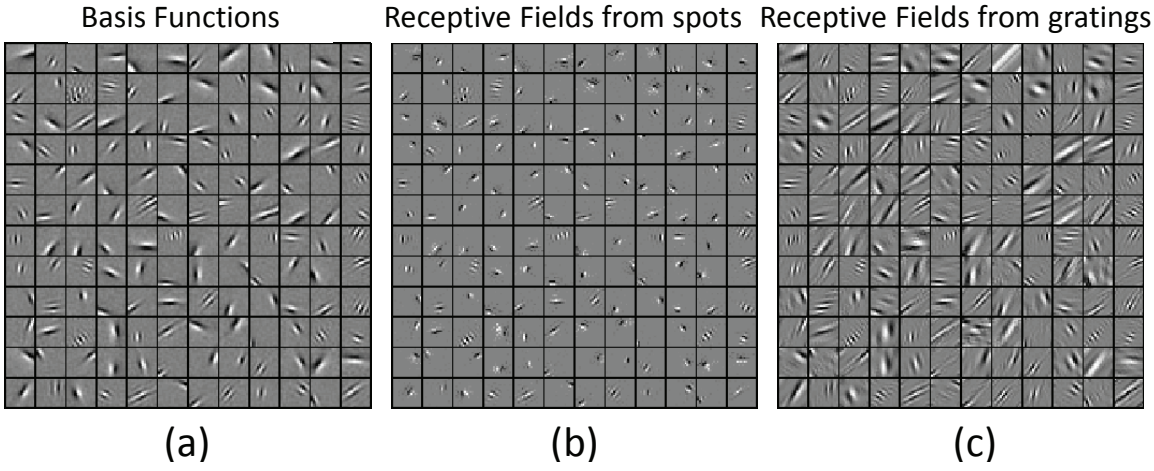


Figure 6.3: (a) shows the basis functions (feedforward weights) learned using 16 sparse coding network with 2.6× overcompleteness. (b) the receptive fields as response profile mapped using spots, and (c) the receptive fields reconstructed from the inverse Fourier transform of the frequency response(the response to gratings).

Figure 6.4a shows the average angular distance between the receptive fields measured using the different the basis sets and the true optimal stimulus. If we represent the receptive field and the optimal stimulus as vectors in the image state space, then the angular distance between them is the angle between the two vectors. We can see from the plot that the average angle between the basis (the true optimal stimulus) and the receptive fields from spots is 38 degrees, and the average angle between the basis vectors and the receptive fields from gratings is 50 degrees. Figure 6.4b depicts these average angular differences between the receptive fields and the true optimal stimulus in three dimensions. We can clearly see that because of the curvature (hyperselectivity), the estimated receptive fields are far from the true optimal stimulus.

Figure 6.5 shows the average response of neurons when probed with the true optimal stimulus, the receptive fields, and the stimuli in 10000 random

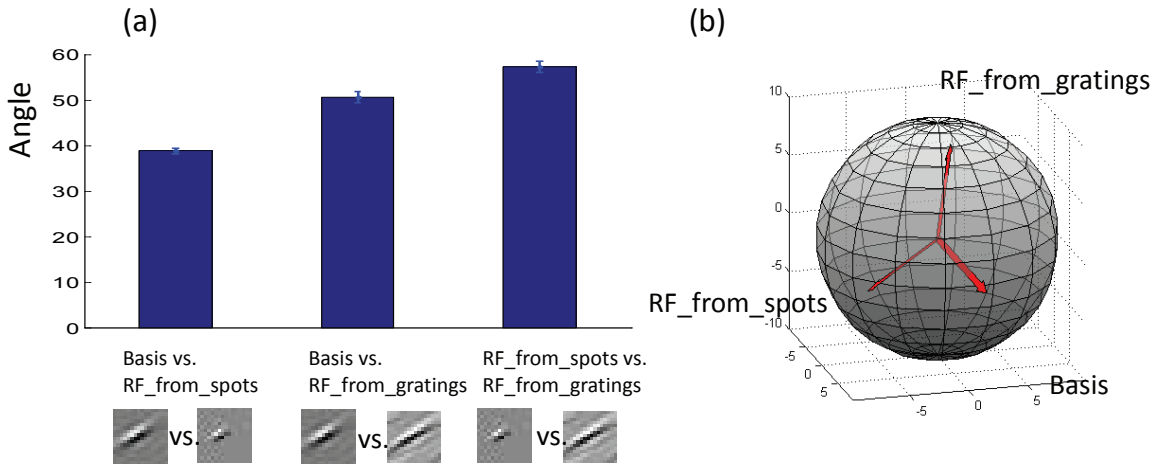


Figure 6.4: (a) shows the average angle between the vectors representing basis function (feedforward weights), receptive field from spots, and the receptive fields from gratings shown in Figure 6.3. (b) shows visually how far the estimation of receptive field is from the basis or the optimal stimulus (S_{max}).

directions and 50 degrees away from the optimal stimulus. Here, we want to emphasize that the feedforward weights of a neuron (the learned basis) is the true optimal stimulus which produces the maximum response. The responses to other stimuli (receptive fields and the 10000 random directions) are normalized such that the response to the true optimal stimulus is 1. Also, the contrasts of the other stimuli (receptive fields and the 10000 random directions) are adjusted such that a linear neuron would produce a response of 1 to these other stimuli. However, from the plot we can see that, irrespective of the contrast adjustment the average response to the other stimuli is less than 1. This implies that there exists a curvature in the neurons' response profiles (curvature in iso-response contours) such that the optimal stimulus (S_{max}) for a neuron is the feedforward weights of the neuron and the neuron is hyperselective to a smaller subspace relative to a linear neuron.

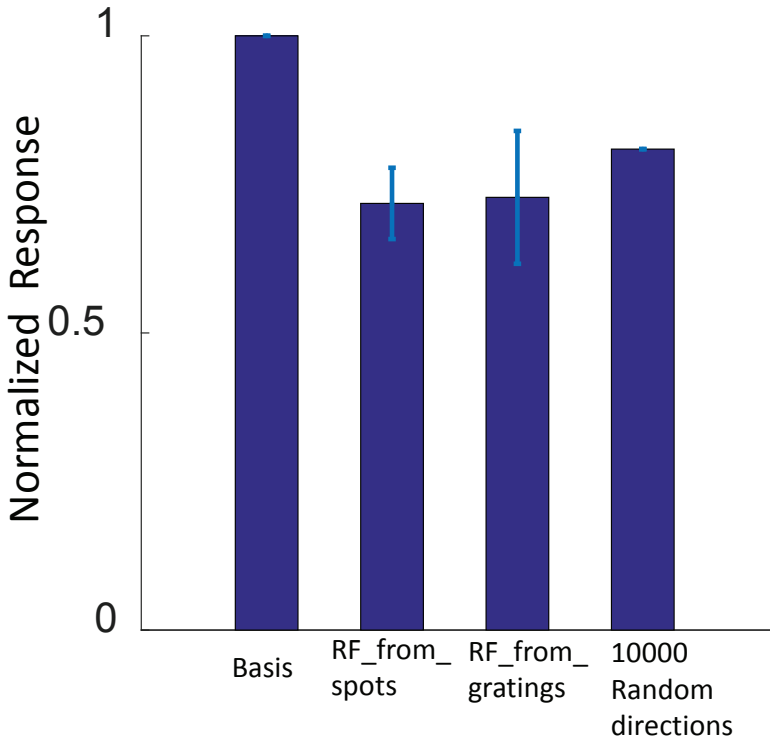


Figure 6.5: The figure shows the relative response of each neuron to stimuli that either match the basis, match the receptive field from spots or match the receptive fields from gratings. The last bar also shows the average response when moving 50 degrees from the basis in 10000 random directions. The results have been normalized such that the responses in all conditions would be 1.0 if the neuron was linear. These results demonstrate that the optimal stimulus for these non-linear neurons is determined by the feed-forward weights(i.e., the basis). This optimal stimulus is not represented by either the receptive fields from spots or the receptive fields from gratings. The hyper-selectivity created by sparse coding significantly reduces the response away from this optimal stimulus.

6.5 The Gabor limit

Marĉelja (1980), introduced Gabor functions (Gabor, 1946) from the telecommunication community to the vision community. He noted the similarities between the V1 simple cells and Gabor functions. Gabor (1946), had argued that his func-

tions are optimized for localizing a signal in both time and frequency. He argued that there is a fundamental limit on knowing simultaneously the width and frequency bandwidth of a signal. This argument was derived from the original uncertainty principle by Heisenberg (1927), where there is a limitation on knowing the exact location and the momentum (energy) of a charged sub-atomic particle. It was argued that the Gabor functions are the ideal trade-off functions in localization of time and frequency. No other function can fall below this trade-off limit (Gabor limit). This optimal trade-off was the reason vision scientists believed that V1 simple cells have the similarities with the Gabor functions.

This Gabor limit is considered as the fundamental limit on the localization factor of neurons. However, this does not hold true when responses depend on nonlinear interactions between neighboring neurons. We believe that many sensory neurons often break this rule. Here we will demonstrate that the sparse coding network breaks this limit. The neurons of the network are more localized in space and the spatial frequency than the Gabor functions.

Figure 6.6 a and b show how the localization factor was computed for the linear feed-forward weights of the neuron and the receptive fields. The localization factor is computed as the product of four bandwidths (two spatial bandwidths and two spatial frequency bandwidths). The two spatial bandwidths (ΔX and ΔY) are estimated by fitting a 2D Gabor function to the spatial map of the neuron (receptive field or the linear feed-forward weights). The remaining two spatial frequency bandwidths (ΔU and ΔV) are estimated by fitting a 2D Gaussian to frequency response (response to gratings). Using these four bandwidth measures, the localization factor is computed as:

$$LocalizationFactor = \Delta X * \Delta Y * \Delta U * \Delta V \tag{6.5}$$

Gabor (1946), argued that there is a fundamental limit on the localization. This was extended by Daugman (1990) to 2D where the localization factor is $1/4\pi^2$. We refer to this as the Gabor limit, and it is argued that no function can fall below this limit.

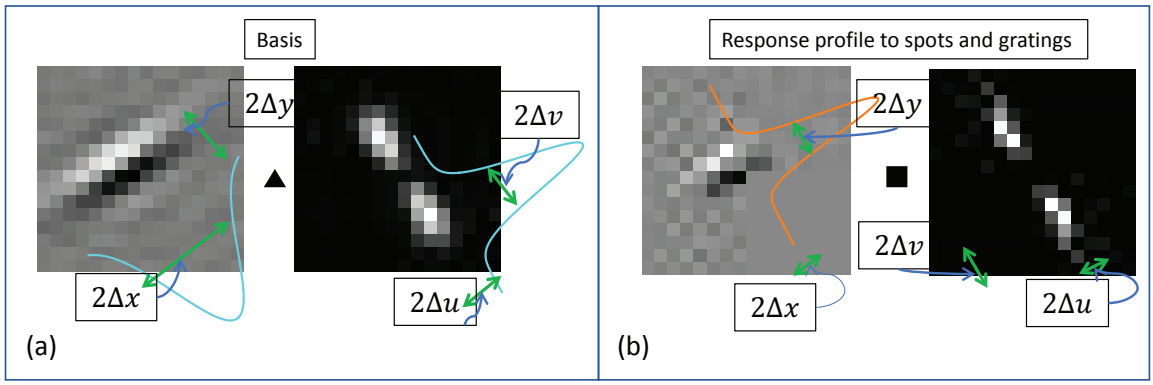


Figure 6.6: The figure demonstrates how the spatial and frequency bandwidths are estimated for each basis function and the estimated receptive fields from spots and gratings. For each neuron in the network we measured the width in space(ΔX and ΔY) and the width in frequency(ΔU and ΔV). For a Gabor function, the product of these widths($\Delta X * \Delta Y * \Delta U * \Delta V$) will be $1/4\pi^2$. We call this product the localization factor and we plot the results for each neuron in the network in Figure 6.7.

The localization factor was computed for every neuron in 2.6 and 4.9 times overcomplete sparse coding networks. For each neuron in the network, two localization factors were computed. One localization factor was computed assuming that the neurons responded linearly. Figure 6.6a shows how the four bandwidths were measured for a linear neuron. The spatial bandwidths (ΔX and ΔY) were computed using the feed-forward weights of the neuron, and the frequency bandwidths (ΔU and ΔV) were computed using the linear frequency response profile (for a linear neuron this is equivalent to the Fourier transform of the 2D spatial map of the feed-forward weights). The second localization

factor was measured using the responses learned by the overcomplete network (which includes nonlinear interactions). Figure 6.6b shows how the four bandwidths were measured for the nonlinear neuron. The spatial bandwidths (ΔX and ΔY) were computed using the receptive field estimated using the spot stimuli and the frequency bandwidths (ΔU and ΔV) were computed using the frequency response to gratings. The receptive field and the spatial frequency response profile shown in Figure 6.6b is for the same neuron shown in Figure 6.6a. We can observe that, with nonlinear interactions, the neuron is more localized in space and frequency than its linear counterpart.

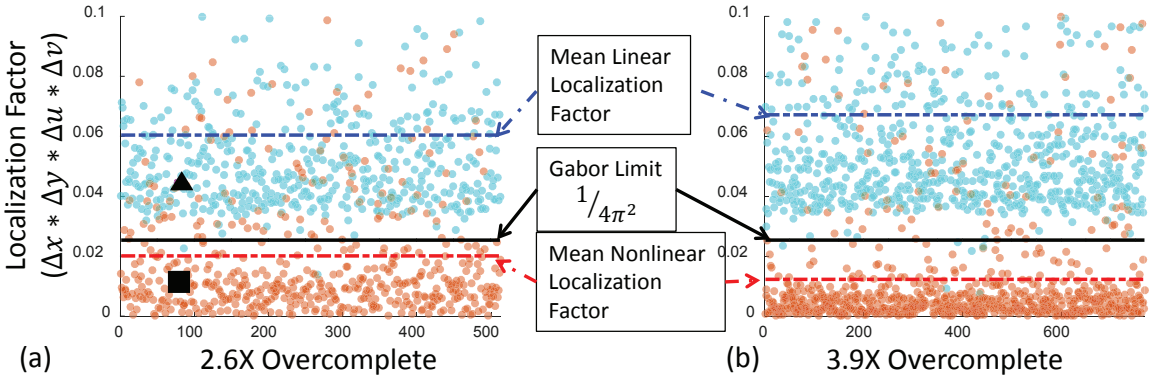


Figure 6.7: The figure shows that the hyperselectivity can produce neurons that are more localized that the predicted Gabor limit of $1/4\pi^2$ (represented by a solid black line). (a) and (b) show the results for a 2.6 times and 3.9 times overcomplete sparse coding network(e.g., Figure 6.3). For each neuron, we plot the localization factor for the feedforward basis(cyan) and the localization factor following nonlinear interactions that produce hyper-selectivity(orange). The dotted lines show the mean localization factor for the linear and the nonlinear conditions. The triangle and square in a) represent the neuron as depicted in Figure 6.6(a) and (b).

Figure 6.7a and b show the scatter plot of localization factors in 2.6 times and 3.9 times overcomplete sparse coding networks. The x-axis in each plot repre-

sents the neuron number in the network. The blue dots represent the localization factor for linear neurons. The dashed blue line shows the mean localization factor for linear neurons. The black solid line represents the Gabor limit. Any linear neuron cannot fall below the line representing the Gabor limit. If the neurons in the network were perfect Gabor functions, then all the blue points would have fallen on the black line. The orange dots represent the localization factor for each neuron which goes through the nonlinear interaction of the network. We can see that for most of the neurons these dots fall below the Gabor limit. Also, the mean localization factor (orange dashed line) falls further below with the increase in overcompleteness.

This result shows that the nonlinear interaction between the neurons can make a neuron break the Gabor limit by becoming hyperselective in multiple domains simultaneously (space and frequency). However, we do not believe the goal of the network is to become hyperselective or localized in these two domains (as thought traditionally). Rather, we believe that the goal of the network is to produce efficient, sparse, and distributed representation of the natural scene environment. In order to achieve this goal, one needs to build overcomplete networks which introduce a lot of redundancies. The mechanisms to get rid of these redundancies produce hyperselectivity as a byproduct.

CHAPTER 7

THE EFFECT OF THE LEARNING RULE

A variety of algorithms have been developed to find efficient representations of natural scene data (for example, the sparse coding network (Olshausen and Field, 1996), Independent Component Analysis (Bell and Sejnowski, 1997), Karklin-Lewicki hierarchical model (Karklin and Lewicki, 2005)). When applied to natural scene data, these algorithms learn receptive fields which have properties similar to the receptive fields of the neurons in the visual pathway. These algorithms are not only limited to visual neurons, as they can also learn efficient representations for other sensory data. However, as discussed in previous chapters, researchers have focused primarily on the linear features learned from these algorithms. These representations are usually two-dimensional receptive fields, which indicate only the optimal direction of a neuron in image state space along which it responds maximally. However, as discussed in previous chapters and other papers Golden et al. (2016); Vilankar and Field (2017), these representations also have an interesting nonlinear response geometry, which yields insight into their selective and invariant responses and could describe a wide family of nonlinearities. In this chapter, we will investigate the role of different learning rules or cost functions, of the sparse coding network on the response geometry and the interaction between the neurons. We will demonstrate how different learning rules affect efficient representation through the changes in response geometry.

The sparse coding network learns to find the directions in the image state space that align with the causes of the data. Usually, these directions are represented as two-dimensional receptive field maps and can be described by Gabor

functions. As the network becomes overcomplete, these representations also develop receptive fields that do not resemble Gabor functions (for example, 10 times overcomplete networks learn receptive fields that are blobs and gratings). As discussed earlier, most researchers focus on the learned optimal directions, or receptive fields, of the neurons in the network. However, we focus here on the response geometry and the interactions of the neurons, as they provide a deeper understanding of the non-linear behavior of the neurons. In previous chapters, we discussed how an overcomplete network produces curvature (hyperselectivity) in iso-response surfaces to handle the non-orthogonal overlap between neurons in the image state space. This curvature in the iso-response surfaces is caused by the sparsifying cost function in the energy function of the sparse coding network (see Equation 5.2 and 5.3).

The cost function is the term which penalizes the network when its responses to the natural scene data are not sparse. As the network becomes overcomplete, this cost function produces non-linear inhibition to reduce the redundancy caused by non-orthogonal neighboring neurons. There are multiple options for the cost function, with each variation producing a different learning rule for the network by which it learns an efficient representation. The popular choices of the cost functions are 'absolute' ($abs(x)$), 'Cauchy' ($log(1+x^2)$), and 'exponential' ($-exp(-x^2)$). Figure 7.1 shows the cost imposed on the network as a function of a neuron's response magnitude. It is widely believed that the choices of cost function or learning rule does not have any significant impact on the basis functions or receptive fields that are learned (Körding et al., 2003). Körding et al. (2003) demonstrated that networks with different cost functions learn basis functions with similar receptive fields, but they also have some qualitative differences. Here, we argue that these different cost functions produce interesting

differences in the response geometry which affects the overall objective (energy functions) in interesting ways. We will demonstrate that different cost functions produce different non-linear response behavior (hyperselectivity) in the image state space. This hyperselectivity causes different patterns (tiling) in the interaction between neighboring neurons which in turn produces interesting local regions in the state space with different reconstruction error (the other objective of the network in addition to sparsity).

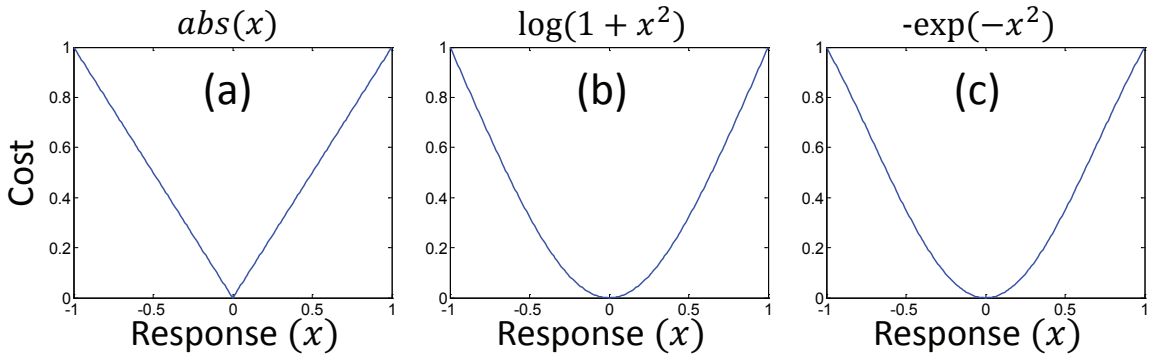


Figure 7.1: The figure shows the three popular choices of the cost function in sparse coding network. The plot shows the magnitude of the cost penalty imposed on the network as the response magnitude increases. (a) shows the 'absolute' cost function ($abs(x)$), (b) shows the 'Cauchy' cost function ($log(1+x^2)$), and (c) shows the 'exponential' cost function ($-exp(-x^2)$).

7.1 Results

In this chapter, we will analyze the role of cost functions in the sparse coding network. This analysis will demonstrate how different cost functions affect the nonlinearity (curvature) of the response of a neuron in the sparse coding network. We will evaluate how different nonlinear behavior due to cost functions

and overcompleteness affect the efficiency of the network in representing the image state space.

For the analysis, we used sparse coding networks on data that was 64-dimensional (8×8 natural scene data), 3-dimensional, and 2-dimensional. To observe the effect of different cost functions, we trained three 64-dimensional networks with three different cost functions: 'absolute', 'Cauchy', and 'exponential'. For each network, all the other variables (eta, lambda, input dimensionality, output dimensionality, etc.) were kept the same. Each network was initialized with the same random initialization. The only changing variable in these networks was the cost function. The basis vectors (Φ) and the responses of the neurons in the network(a) were learned using gradient descent on the energy function (Equation 5.4 and 5.8).

To visualize the effect of the learning rules on the response geometry and the interactions between the neurons, we used lower dimensional (2D and 3D) sparse coding networks. In these networks, the basis functions were not learned. The directions of the basis functions in the image state space were uniformly distributed and fixed. The responses of the neurons represented by the basis functions were learned using the sparse coding network of different cost functions. For Figure 7.2 and 7.3 the directions of the basis vectors were hand-picked such that the angle between the adjacent vectors was 70 degrees. For Figure 7.4-7.5 and 7.7-7.9, the directions of the 14 basis vectors were hand-picked such that the angle between adjacent basis vectors was same. In these figures the 3D sparse coding network was probed using points which lie on a sphere in that 3D image state space.

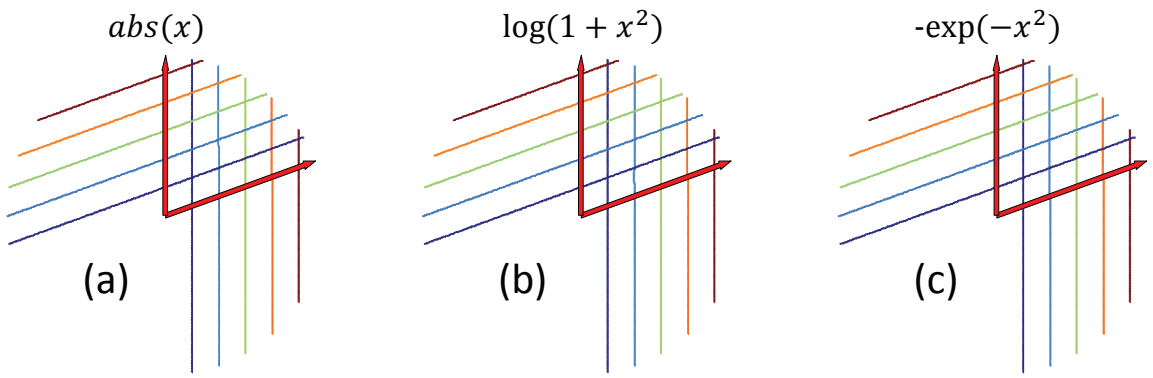


Figure 7.2: The effect of the three cost functions on the iso-response contours of the vertical neuron in 2D image state space. In this image state space, only two neurons exist which are 60 degrees apart and are represented by the two red vectors. One can note that the iso-response contours tilt instead of warping around the vector. The warping of the iso-response contours can be achieved in the 2D state space if there is a third non-orthogonal vector on the left of the vertical neuron. Also, there is not much effect of the different cost functions in 2D image state space.

7.1.1 The effect of learning on the iso-response contours in 2D subspaces

First, we will observe the effect of different cost functions of the sparse coding network on the iso-response contours of a neuron. For this analysis, I selected two neurons which were represented by two vectors separated by 60 degrees in 2D state space (or in a 2D subspace of high-dimensional image state space). Figure 7.2 shows the effect of the three cost functions on the iso-response contours in the 2D sparse coding network. The figure shows the iso-response contours of the vertical vector. Figure 7.3 shows the iso-response contour in the 2D subspace defined by the vectors separated by 60 degrees in 64D image state space. The first difference we notice is that the 2D sparse coding network tilts the iso-

response contours whereas the 64D sparse coding network learns to warp the iso-response contours around the vector. Currently, we believe that the neurons from other dimensions are influencing the response of the neurons. Figure 7.3 clearly shows that the curvature depends on the angle (as discussed in the previous chapters). The three cost functions can be seen to produce small but different forms of curvature in 2D. We believe that in order to see substantial difference in the curvature we need to probe the network with a larger proportion of the image state space. The results of Figure 7.3 were obtained by probing a 2D subspace in 64D image state space. In order to probe a larger proportion of the image state space, we used 3D sparse coding network with 14 basis vectors uniformly distributed in the state space.

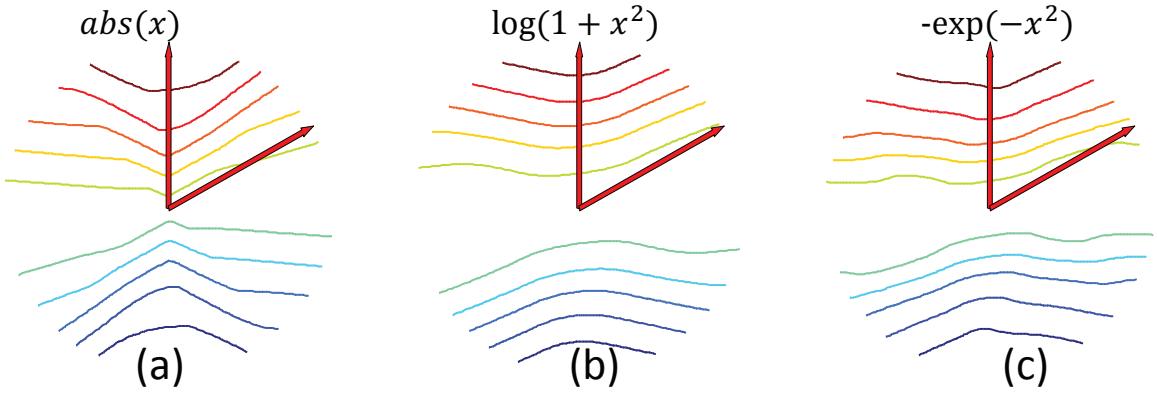


Figure 7.3: The effect of the three cost functions on the iso-response contours of the vertical neuron in a 2D subspace of high dimensional (8) sparse coding network. The network is 2.6 times overcomplete with 128 neurons. We selected a pair of neurons which are 60 degrees apart in the image state space. The network is probed only with the data points in the 2D subspace defined by the vectors. One can note that the network learns to warp (with exo-origin curvature) the iso-response contours and the curvature depend on the angle between the vectors. The three cost functions (a,b, and c) appears to have a small effect on the iso-response contours.

7.1.2 The effect of the learning rule on selectivity

To observe a significant effect of the learning rule we used toy data from a 3D sparse coding network. As mentioned before the network had 14 basis vectors uniformly distributed in the state space. The directions of the vectors were not learned. However the network was probed with 3D data points, and the response was learned using gradient descent on the energy function (See Equation 5.4) which includes the cost function. The network was probed with a specific data set that lies on the sphere with unit radius in the state space. Figure 7.4 shows the response of the green vector as a heat map on the sphere. The color represents the response magnitude to each data point on the sphere. The red color represents a high response, and the dark blue color represents a zero or a small response. Figure 7.4a shows the response of the positively responding linear neuron. Such a neuron will only respond to a half of the sphere, for the other half it will respond as zero. Figure 7.4b,c, and d show the learned response of the neuron using the three cost functions. First of all, we can notice that the neuron from the sparse coding network (for all of the cost functions) responds only to a small region on the sphere. This indicates that because of the curvature of the iso-response surfaces, the neuron has become hyperselective and responds to a small region. However, the interesting thing to note is that because of these different sparsification learning rules or cost functions the shape of the hyperselective region for the three cost functions is different. These different shapes of the selective region have multiple implications for how representation of the image state space is shared between the vectors and how this sharing/tessellation leads to a better representation of data (regarding minimum reconstruction error).

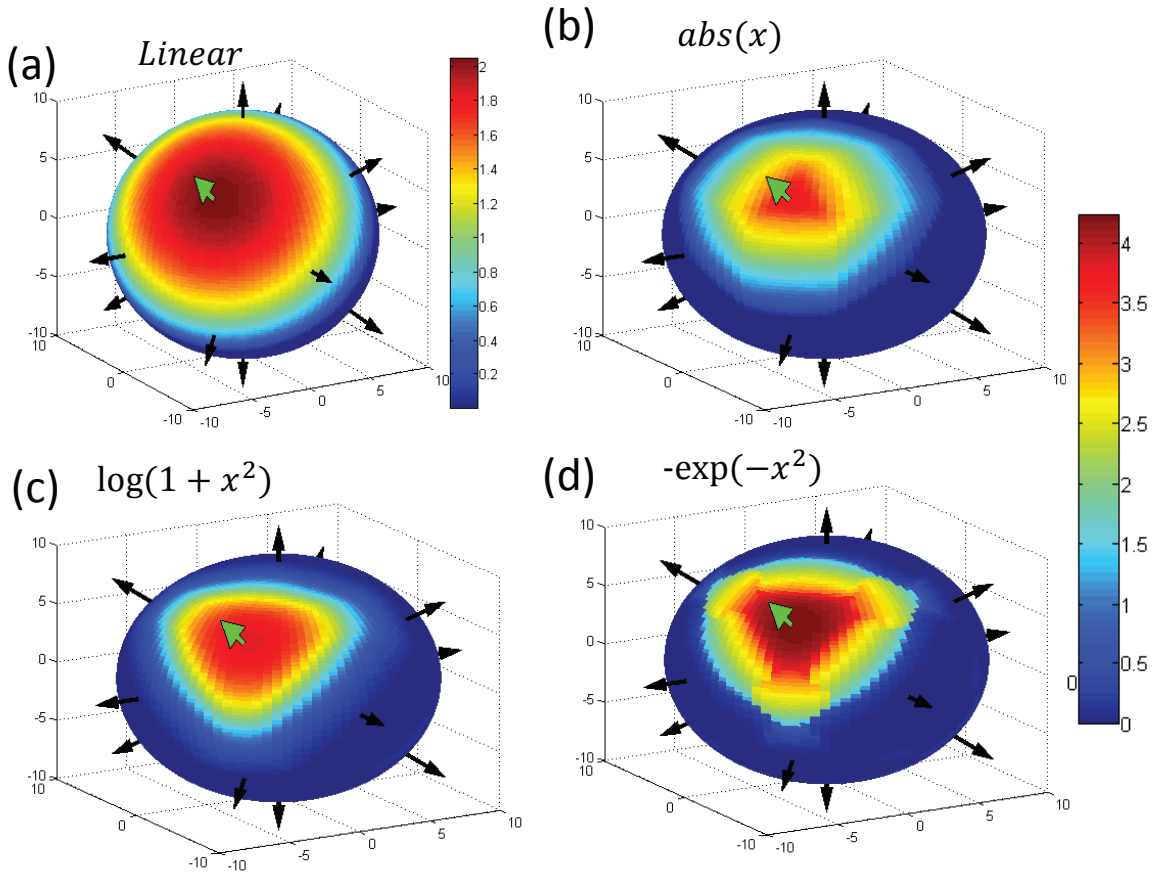


Figure 7.4: The effect of the cost functions on the shape of the hyperselective region of a neuron in 3D sparse coding network with 14 basis vectors. (a) shows the hyperselective region of a linear neuron. (b), (c), and (d) show the hyperselective region of a non-linear neuron with 'absolute', 'Cauchy', and 'exponential' cost functions respectively.

7.1.3 The effect of learning rule on tiling

In Figure 7.4, we observed the effect of different learning rules on the shape of hyper-selective regions in the image state space. In the previous chapters, we discussed how the neighboring neurons (the angle between neighboring vectors) affect the selectivity of a neuron, producing a hyperselective region due to curvature. Now, we will analyze how these hyperselective neurons interact

and share the state space between them. Figure 7.5 shows how space is shared by different vectors/ neurons. For this analysis, the response of all 14 vectors was computed for each point on the sphere. The responses at each point were then sorted in descending order. In Figure 7.5 each color represents a unique combination of three neurons which produced the maximum response at each data point. We can observe that usually, the neighboring data points on the sphere have the same colors which produce a tiling of data points on the sphere. Each tile represents a unique combination of three neurons which are responding maximally in that region of the state space. Figure 7.5a-d, show different the tiling arrangements due to different cost functions used for learning the responses in the sparse coding network.

It is possible to have more than or less than three neurons responding in the region. However, if we observe the average response in the seven maximally responding neurons at each position, we can see that most of the response energy is in the top three neurons, with the exception of the linear model. This is expected because ideally only three neurons should respond in a three-dimensional state space with an overcomplete number of neurons. Ideally, in an n-dimensional image state space, no more than n vectors should be responding to any given image. Similarly, in 3-dimensional image state space, only three of the vectors should respond for any given image. Figure 7.6a-d show the average in the top seven neurons. For the linear model, the fall-off in the average response energy is linear as expected. However, for the sparse coding models with different learning rules, most of the energy is the top three neurons. Furthermore, energy in the first vector compared (Figure 7.6b-d) to the first vector from the linear model (Figure 7.6a) indicate that the sparse coding network tries to represent the data points with a fewer number of vectors than the dimension-

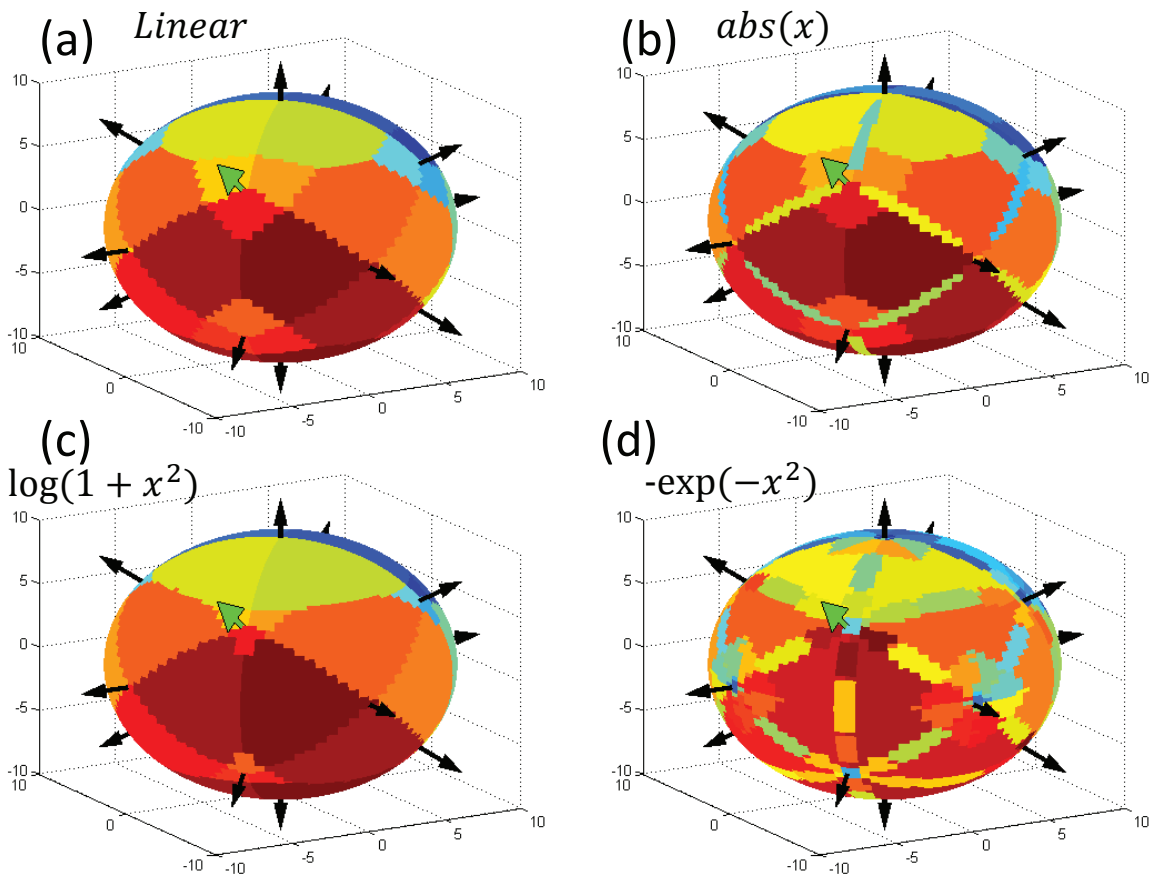


Figure 7.5: The effect of the cost functions on the sharing of the image state space between neurons. Each color tile represents a unique combination of three neurons which produced the maximum response. (a) Shows the sharing between linear neurons. (b), (c), and (d) show the sharing between neurons of the sparse coding network with 'absolute', 'Cauchy', and 'exponential' cost functions respectively.

ality of the data.

Figure 7.7a-d shows the cumulative response energy in 4th, 5th, 6th, and 7th neuron as heat map at each data point on the 3D sphere. Figure 7.7a shows a lot of cumulative response energy all over the sphere, and for the sparse coding network models, the cumulative response energy is much less. However, one should note the interesting differences in this figure. For the 'absolute' cost

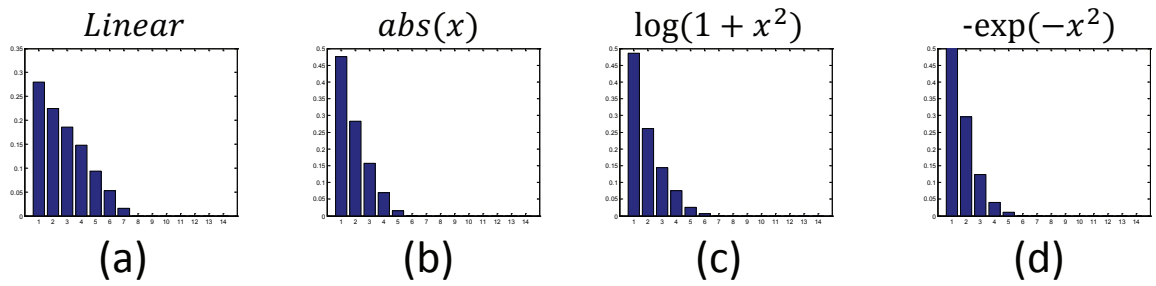


Figure 7.6: The figure shows the average response energy in seven most responding neurons at each data point on the sphere. (a) shows the response energy for linear neurons. (b), (c), and (d) show the response energy for neurons of the sparse coding network with 'absolute', 'Cauchy', and 'exponential' cost functions respectively.

function, there are regions near the vectors and between the vectors which show relatively higher cumulative energy than the other regions. Similar behavior can be observed for the 'Cauchy' cost functions. However, for the 'exponential' cost function the regions near the vector show much less energy. If we assume that the sparse coding network learns the direction of the causes of data in the image state space and points the neurons in those directions then, the network should represent those directions with sparseness producing little or no response energy in the 4th, 5th, 6th, and 7th neuron. In an efficient solution, all the energy should be in the top *n* neurons (where n is the dimensionality of the state space), with no energy in the other neurons.

7.1.4 Effect of learning rule on tiling and reconstruction

So far, we have seen the effect of learning rule on the geometry of responses, the sharing of space by the vectors, and the ability to effectively produce a sparse solution. Now we will see how all of these affect the ability to reconstruct the

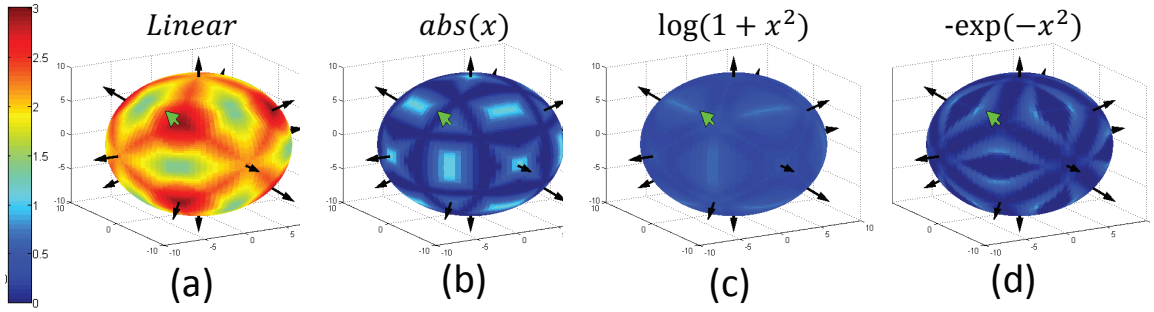


Figure 7.7: The figure shows the cumulative response energy in 4th, 5th, 6th, and 7th neuron (excluding the top three maximally responding neurons).

input data. Reconstruction is one of the important constraints on the sparse coding network. The reconstruction error is measured $(I - \phi A)^2$. Figure 7.8a-c shows the accuracy of the sparse coding network in reconstructing the input data as a heat map on the sphere. This figure gives us a view of how the sparse coding network is performing in reconstructing the data at different regions of the input data space. Figure 7.8a-c show the reconstruction error for the 'absolute', 'Cauchy', and 'exponential' cost function. We can notice the small red patches near the vector for the 'absolute' cost function; this implies that 'absolute' cost functions is relatively the worst for reconstructing the input near the vector than at other regions. However, 'Cauchy' and 'exponential' are dark blue near the vectors suggesting good reconstruction. Again, if we assume that the sparse coding network learns the direction of the causes of data in the image state space and points the neurons in those directions, then the network should represent those directions with minimum reconstruction error. Also, the 'Cauchy' cost function has red patches (with high reconstruction error) in the region between the vectors whereas the 'exponential' cost function has very small regions with high reconstruction error.

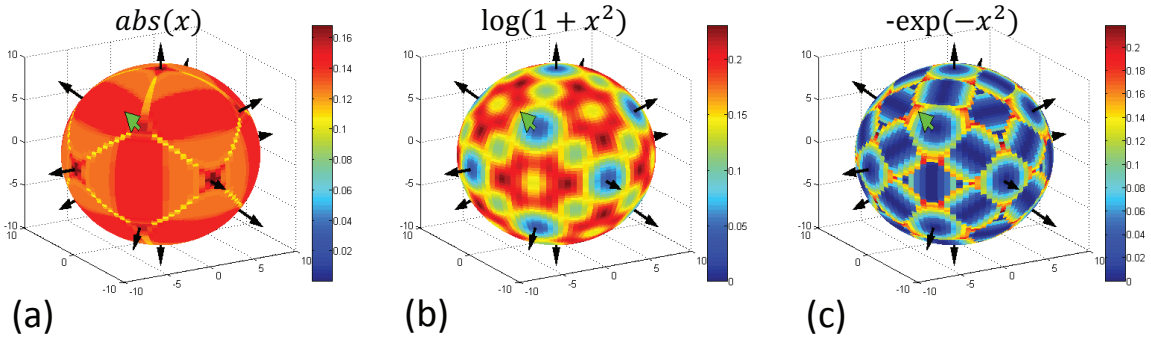


Figure 7.8: The figure shows reconstruction error as a heat map on the sphere. (a), (b), and (c) show the reconstructing error for sparse coding networks with 'absolute', 'Cauchy', and 'exponential' cost functions respectively.

Figure 7.9a-c show the reconstruction error for the 'absolute', 'Cauchy', and 'exponential' cost function on a relative scale. From the figure, we can observe that the 'absolute' cost function produces the maximum reconstruction error and the 'exponential' cost function produces the least reconstruction error. From the results of Figure 7.7,7.8, and 7.9, it appears that the 'exponential' function is more efficient in representing the data regarding both the reconstruction error and the sparsity.

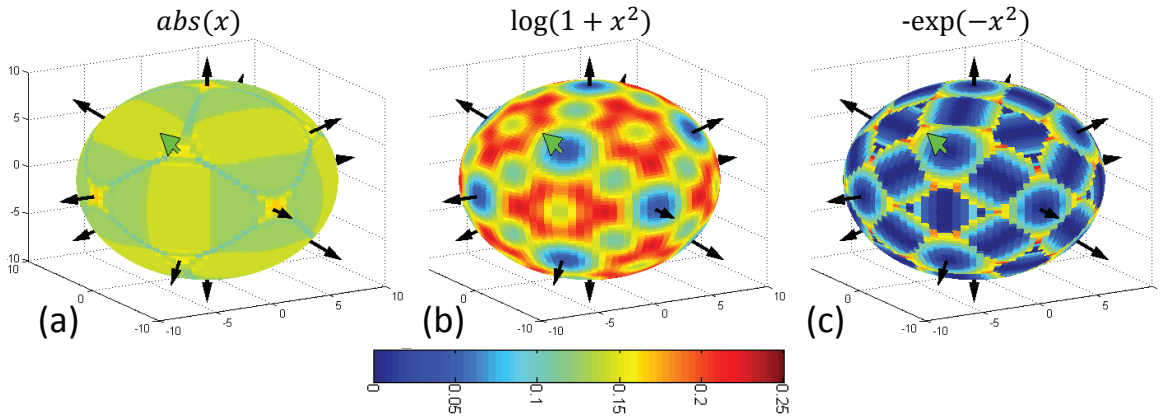


Figure 7.9: The figure shows reconstruction error as a heat map on the sphere. This figure uses same heat map scale for all three figures. (a), (b), and (c) show the reconstruction error in sparse coding networks with 'absolute', 'Cauchy', and 'exponential' cost functions respectively.

CHAPTER 8

CONCLUSION AND FUTURE WORK

In this dissertation, I attempted to reveal some of the efficient information processing mechanisms of the mammalian visual system. The work that I presented argues that that the neurons in the visual system are not simple feature detectors. The linear filters estimated from the receptive fields are not complete descriptions, and the neurons perform functions above and beyond simple feature detection. For this work, I conducted psychophysical experiments to capture the statistics of different categories of edges in the natural images and developed theoretical framework regarding a possible mechanism to segregate the different causes of an edge in the early visual system. I described a geometrical framework that we developed to describe a wide family of nonlinearities observed in V1. Further, I described how the geometrical visualization of nonlinear responses in overcomplete sparse coding networks provides deeper insights into efficient coding mechanisms.

In Chapter 2, I began by exploring the statistics of different categories of edges in the natural scene environment and demonstrated that different cells in the early visual system could potentially perform different computations to start building probabilistic inference about the edge categories. We found that approximately 50% of the edges in natural images were labeled as occlusion edges, and that the local statistics of edges contain significant information about the edge category. For example, the local contrast of an occlusion edge was significantly higher than the local contrast of a nonocclusion edge. We infer (by developing a maximum likelihood classifier) that the early visual system could potentially use these statistics to segregate figure from ground, which is a cru-

cial step for object recognition by the later stages of the visual system. In the future, I would like to work further to see if these different statistics could produce different classes of V1 receptive fields. The psychophysical experiments have already observed multiple (at least 30) classes of retinal ganglion cells (Sanes and Masland, 2015). I would like to explore theoretically if different neural networks (e.g., the sparse coding network, the Karklin-Lewicki hierarchical network, and deeper networks) could learn different neurons (with respect to optimal stimulus and hyperselectivity) from the local statistics of edges.

In Chapters 3 and 4, I described the argument we made in Golden et al. (2016) that the early nonlinearities in the visual system can be described by a simple warping curvature in the iso-response manifolds of neurons. We described two forms of warping curvature: exo-origin curvature and endo-origin curvature. We argued that exo-origin curvature describes selective nonlinearities such as end-stopping, cross-orientation inhibition and non-classical receptive field effects. Similarly, we argued that endo-origin curvature describes invariant/tolerant nonlinearities such as complex cells. In Chapter 4, we explored four different models which generate the required warping curvature. In Chapter 5 we argue that exo-origin curvature in an overcomplete sparse coding network produces an efficient sparse representation. We show that the curvature gets rid of the redundancy in the representation produced by the overcomplete set of neurons. We argue that in an n-dimensional system, no more than n neurons should be active. In Golden et al. (2016), we describe a fan equation model which guarantees that in two-dimensional image state space no more than two vectors are active for any given data point, even though the number of encoding vectors is overcomplete (more than two). The third model that we describe is the gain control model, which was not designed to produce curvature, but rather to

produce the gain control response through divisive normalization. Finally, we explore how the cascaded linear non-linear model also produces curvature. All these models have subtle differences (discussed in Chapter 4) but all show curvature in iso-response manifolds. Currently, we do not believe there is sufficient physiological evidence to distinguish between these models.

In Chapter 5, we took a closer look at the curvature in an overcomplete sparse coding network. I demonstrated that the receptive fields (feedforward weights) learned using sparse coding networks of different overcompleteness look very similar and can be well described by Gabor functions. However, the mean amount of exo-origin curvature produced in the iso-response curves varied as a function of overcompleteness (see Figure 5.7c). We quantified the curvature by estimating a parabolic fit to the iso-response curves. We demonstrated that the curvature depends on the angle between neighboring vectors and the overcompleteness. As shown in Figure 5.6, the smaller the angle between neurons, the greater the curvature. Similarly, the mean curvature increased with increasing overcompleteness of the network. We also found that curvature produced by the sparse coding network in high-dimensional image state space was less than the curvature produced by the fan equation model in two-dimensional state space. Currently, we are working towards to develop Fan equation model in higher dimensional state space.

The method to quantify the curvature that I described estimates only the curvature in a 2D subregion defined by two neighboring neurons in the high-dimensional state space. Golden (2015) developed a method to measure the curvature in high-dimensional state space by applying the tools from differential geometry to iso-response manifolds. This method allows us to measure

and compare the curvature in different network models. For example, Golden (2015) demonstrated that in the overcomplete sparse coding network, neurons produce positive curvature of varying magnitude, which implies exo-origin curvature. Whereas a neuron from the second layer of Karklin & Lewicki network (Karklin and Lewicki, 2005) produce positive as well as negative curvature, implying both exo-origin and endo-origin curvature. This means that neurons from the second layer of Karklin & Lewicki network show simultaneous selective and invariant/tolerant nonlinearities. There are principle dimensions where the neuron behaves as selectively, and there are dimensions where the same neuron respond invariably.

Over the last few years, deep learning algorithms (LeCun et al., 2015; Szegedy et al., 2015; He et al., 2016) have seen tremendous success in the machine learning community and allowed for the development of commercial products. These deep learning techniques have been applied in a variety of areas such as computer vision, speech recognition, and self-driving automobiles. The success of deep learning networks has motivated researchers to understand the inner working of the neurons in the network. One approach to understanding the inner working is through the technique called "Deep Visualization" (e.g., Zeiler and Fergus (2014); Yosinski et al. (2015); Mahendran and Vedaldi (2016); Nguyen et al. (2016). In this technique, a synthetic image is created that maximally activates a neuron. However, as discussed before we firmly believe that the optimal stimulus that activates a neuron maximally is not the complete description, there are more interesting aspects of a neuron in its response geometry. We are currently investigating the techniques to measure the curvature in the response of neurons in deep networks. It is believed that for any successful object recognition model it is necessary to learn the low di-

mensional object manifolds that exist in the high-dimensional image state space (Edelman, 1999; Field and Wu, 2004; DiCarlo and Cox, 2007; Golden et al., 2016). DiCarlo and Cox (2007) believe that the projections of these manifolds are ugly and highly convoluted, whereas Field and Wu (2004) believe that these manifolds are not ugly but have a systematic structure. The warping nonlinear neurons at various stages of the visual system untangle these manifolds such that they are linearly separable by the neurons of higher stages (e.g., Inferotemporal Cortex). Recently, algorithms for object recognition using deep learning have demonstrated accuracy equal to human performance, or even exceeding it in some specific cases. I believe that quantifying the curvature in higher dimensional image state space along with the knowledge of the optimal stimulus of a neuron would enable us to understand how the deep learning networks of object recognition are able to separate out the tangled low dimensional object manifolds.

In Chapter 6, we demonstrated the effect of curvature on the response selectivity of a neuron around its optimal stimulus. We termed this selectivity as "hyperselectivity". Due to the hyperselectivity, we showed that it is possible to have a paradoxical neuron which is narrowly tuned to a broadband stimulus. A neuron which has an optimal stimulus which is broadband (e.g., broad in spatial frequency or orientation), can appear to be narrowly tuned. Small deviations from the optimal stimulus (e.g., changes to orientations) can cause the neuron to shut off. However, the optimal stimulus remains a broadband stimulus. We demonstrated that because of hyperselectivity, the estimates of receptive fields could be misleading. The estimates of receptive fields depend on the orthonormal basis set used to probe the neurons. In the sparse coding network, the estimated receptive fields using a grating basis set were different

from the estimated receptive fields using a spot basis set. Also, the estimated receptive fields from the two different orthonormal basis set were different from the actual optimal stimulus (the feedforward weights). I believe that to estimate the true optimal stimulus or the true receptive field, we need to develop new techniques of physiology that can densely probe a neuron around its optimal stimulus rather that probing with an orthonormal basis. We believe that the method that uses the spike triggered covariance (e.g., Schwartz et al. (2002); Rust et al. (2004); Vintch et al. (2015)) can find the relevant subspaces where the nonlinear interaction between neighboring neurons occur. This approach finds the inhibitory (hyperselective) and the excitatory (invariant) dimensions. By probing the neurons densely in the relevant inhibitory subspaces it might be possible to see the accurate nature of the iso-response contours.

Finally, in Chapter 7, I show the interaction between the neurons of the sparse coding network and the effect of the cost function on hyperselectivity. From Chapter 2 to Chapter 6 we focused on the effects of curvature on the response geometry of individual neurons. However, the interaction between neurons could also provide deeper insights into the information processing mechanisms of the network. We observed that in the sparse coding network different learning rules produce different shapes of hyperselective region. The hyperselective region of each neuron interacts with the hyperselective regions of other neurons, which we visualized by exploring the tiling of the image state space in low dimensions and low-dimensional subspaces (see Figure 7.5). The tiles in the state space show how different neurons share different regions of the state space. The tiles representing the shared space, in turn, produce regions of space that are reconstructed with varying amount of error.

It is known that the natural images are restricted to small regions/directions of the image state space (Kersten, 1987; Field, 1987). It is important that any efficient representation mechanism should identify the restricted region and share the region with other units in a way that has the lowest possible reconstruction error. I believe that the visualization of interaction and space sharing between the units would provide deeper insights into the efficient representation mechanism of artificial or biological networks. In Chapter 7, I show visualizations of neuron interaction in a three-dimensional state space with toy data. However, the natural image state space is high dimensional, and it is rare to find more than three neurons in a 3D subspace. In the future, I would like to develop better visualizations and measures of interaction between neurons in higher-dimensional image state space.

BIBLIOGRAPHY

Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *JOSA A*, 2(2):284–299.

Albrecht, D. G. and Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, 7:531–546.

Albrecht, D. G., Geisler, W. S., and Crane, A. M. (2003). Nonlinear properties of visual cortex neurons: Temporal dynamics, stimulus selectivity, neural performance. In Chalupa, L. and Werner, J., editors, *The Visual Neurosciences*, pages 825–837. MIT Press, Boston.

Albrecht, D. G., Geisler, W. S., Frazor, R. A., and Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *Journal of Neurophysiology*, 88(2):888–913.

Albrecht, D. G. and Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *J Neurophysiol*, 48(1):217–237.

Anderson, M. and Pessoa, L. (2011). Quantifying the diversity of neural activations in individual brain regions. In *Proceedings of the Cognitive Science Society*, volume 33.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, 33(4):245–266.

Andrews, B. and Pollen, D. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *The Journal of physiology*, 287:163–176.

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network*, 3:213–251.

Atick, J. J. and Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, 2(3):308–320.

Atick, J. J. and Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural computation*, 4(2):196–210.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, 61(3):183.

Balboa, R. M. and Grzywacz, N. M. (2000). Occlusions and their relationship with the distribution of contrasts in natural images. *Vision Research*, 40(19):2661 – 2669.

Barlow, H. (1979). Three theories of cortical function. In *Developmental Neurobiology of Vision*, pages 1–16. Springer.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *The Journal of physiology*, 119(1):69.

Barlow, H. B. (1961). The coding of sensory messages. *Current problems in animal behaviour*, 331:360.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1(4):371–394.

Barlow, H. B., Blakemore, C., and Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *The Journal of physiology*, 193(2):327.

Bell, A. J. and Sejnowski, T. J. (1997). The independent components of natural scenes are edge filters. *Vision research*, 37(23):3327–3338.

Bonds, A. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual neuroscience*, 2(01):41–55.

Brunswik, E. (1947). Systematic and representative design of psychological experiments. In *Proceedings of the Berkeley symposium on mathematical statistics and probability*, pages 143–202.

Brunswik, E. (1952). The conceptual framework of psychology. *Psychological Bulletin*, 49(6):654–656.

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pages 679–698.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant, J. L., and Rust, N. C. (2005). Do we know what the early visual system does? *The Journal of Neuroscience*, 25(46):10577–10597.

Cavanaugh, J. R., Bair, W., and Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of neurophysiology*, 88(5):2530–2546.

Dan, Y., Atick, J. J., and Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, 16(10):3351–3362.

Daugman, J. G. (1990). An information-theoretic view of analog representation in striate cortex. *Computational neuroscience*, 2:403–423.

David, S. V. and Gallant, J. L. (2005). Predicting neuronal responses during natural vision. *Network: Computation in Neural Systems*, 16(2-3):239–260.

De Valois, K. K. and Tootell, R. (1983). Spatial-frequency-specific inhibition in cat striate cortex cells. *The Journal of Physiology*, 336(1):359–376.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1978). Cortical cells: bar and edge detectors, or spatial frequency filters? In *Frontiers in visual science*, pages 544–556. Springer.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22(5):545–559.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1985). Periodicity of striate-cortex-cell receptive fields. *JOSA A*, 2(7):1115–1123.

Dean, A. and Tolhurst, D. (1983). On the distinctness of simple and complex cells in the visual cortex of the cat. *The Journal of Physiology*, 344(1):305–325.

DeAngelis, G., Robson, J., Ohzawa, I., and Freeman, R. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. *Journal of Neurophysiology*, 68(1):144–163.

DeAngelis, G. C., Ohzawa, I., and Freeman, R. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. ii. linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69(4):1118–1135.

DeValois, R. L. and DeValois, K. K. (1988). *Spatial vision*. Number 14. Oxford University Press.

DiCarlo, J. J. and Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, 11(8):333–341.

DiMattina, C., Fox, S. A., and Lewicki, M. S. (2012). Detecting natural occlusion boundaries using local cues. *Journal of vision*, 12(13).

Dong, D. W. and Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–358.

Edelman, S. (1999). *Representation and recognition in vision*. MIT press.

Elder, J. H. (1999). Are edges incomplete? *Int. J. Comput. Vision*, 34(2-3):97–122.

Elder, J. H., Beniaminov, D., and Pintilie, G. (1999). Edge classification in natural images (abstract). *Investigative opthalmology & visual science*, 40:s357. Presentation available at `http://elderlab.yorku.ca/~elder/?page=pub&lb=lbNone`.

Enroth-Cugell, C. and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of physiology*, 187(3):517–552.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394.

Field, D. J. (1994). What is the goal of sensory coding? *Neural computation*, 6(4):559–601.

Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Res.*, 33:173–193.

Field, D. J. and Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proceedings of the Royal society of London. Series B. Biological sciences*, 228(1253):379–400.

Field, D. J. and Wu, M. (2004). An attempt towards a unified account of nonlinearities in visual neurons. *Journal of vision*, 4(8):283.

Fowlkes, C. C., Martin, D. R., and Malik, J. (2007). Local figureground cues are valid for natural images. *Journal of Vision*, 7(8):2.

Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441.

Gardner, J. L., Anzai, A., Ohzawa, I., and Freeman, R. D. (1999). Linear and nonlinear contributions to orientation tuning of simple cells in the cat's striate cortex. *Visual neuroscience*, 16(06):1115–1121.

Geisler, W., Perry, J., Super, B., and Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision research*, 41(6):711–724.

Gibson, J. J. (1950). The perception of the visual world.

Golden, J. R. (2015). *A unified approch to the non-linearities of visual neurons: the curved geometry of neural response surfaces.* PhD thesis, Cornell University.

Golden, J. R., Vilankar, K. P., Wu, M. C., and Field, D. J. (2016). Conjectures regarding the nonlinear geometry of visual neurons. *Vision research*, 120:74–92.

Hartline, H. K., Wagner, H. G., and Ratliff, F. (1956). Inhibition in the eye of limulus. *The Journal of general physiology*, 39(5):651–673.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–197.

Heisenberg, W. (1927). Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43(3-4):172–198.

Heitger, F., von der Heydt, R., and Kubler, O. (1994). A computational model of neural contour processing: Figure-ground segregation and illusory contours. In *Proceedings of IEEE on Perception to Action*, pages 181–192. IEEE.

Hoiem, D., Efros, A. A., and Hebert, M. (2011). Recovering occlusion boundaries from an image. *International Journal of Computer Vision*, 91(3):328–346.

Howe, C. Q. and Purves, D. (2002). Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academy of Sciences*, 99(20):13184–13188.

Huang, J., Lee, A. B., and Mumford, D. (2000). Statistics of range images. In *Proceedings of IEEE on Computer Vision and Pattern Recognition*, volume 1, pages 324–331. IEEE.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160:106–154.

Hubel, D. H. and Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, 28(2):229–289.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195:215–243.

Ing, A. D., Wilson, J. A., and Geisler, W. S. (2010). Region grouping in natural foliage scenes: Image statistics and human performance. *Journal of Vision*, 10(4):10.

Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258.

Kagan, I., Gur, M., and Snodderly, D. M. (2002). Spatial organization of receptive fields of v1 neurons of alert monkeys: comparison with responses to gratings. *Journal of neurophysiology*, 88(5):2557–2574.

Karklin, Y. and Lewicki, M. S. (2005). A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17(2):397–423.

Kersten, D. (1987). Predictability and redundancy of natural images. *J. Opt. Soc. Am. A*, 4:2395–2400.

Körding, K. P., Kayser, C., and König, P. (2003). On the choice of a sparse prior. *Reviews in the Neurosciences*, 14(1-2):53–62.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J Neurophys*, 16:37–68.

Leclerc, Y. G. and Zucker, S. W. (1987). The local structure of image discontinuities in one dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:341–355.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, H., Battle, A., Raina, R., and Ng, A. Y. (2006). Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808.

Leibe, B., Leonardis, A., and Schiele, B. (2008). Robust object detection with interleaved categorization and segmentation. *International journal of computer vision*, 77(1-3):259–289.

Lewicki, M. S. and Olshausen, B. A. (1999). Probabilistic framework for the adaptation and comparison of image codes. *JOSA A*, 16(7):1587–1601.

Lewicki, M. S. and Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural computation*, 12(2):337–365.

Li, W. and Gilbert, C. D. (2002). Global contour saliency and local colinear interactions. *Journal of neurophysiology*, 88(5):2846–2856.

Liu, Y., Cormack, L. K., and Bovik, A. C. (2011). Statistical modeling of 3-d natural scenes with application to bayesian stereopsis. *IEEE Transactions on Image Processing*, 20(9):2515–2530.

Mahendran, A. and Vedaldi, A. (2016). Visualizing deep convolutional neural networks using natural pre-images. *International Journal of Computer Vision*, 120(3):233–255.

Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S., and Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nature neuroscience*, 8(12):1690–1697.

Marĉelja, S. (1980). Mathematical description of the responses of simple cortical cells. *JOSA*, 70(11):1297–1300.

Marr, D. (1982). *Vision: A computational approach*. Freeman.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 207(1167):187–217.

Martin, D. R., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416–423. IEEE.

Martin, D. R., Fowlkes, C. C., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 26:530–549.

Martin, P. R. and Grünert, U. (2004). Ganglion cells in mammalian retinae. *The visual neurosciences*, 1:410–421.

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception*, 33(9):1101–1128.

Mély, D. A. and Serre, T. (2017). Towards a theory of computation in the visual cortex. In *Computational and Cognitive Neuroscience of Vision*, pages 59–84. Springer.

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978a). Receptive field organization of complex cells in the cat's striate cortex. *The Journal of physiology*, 283(1):79–99.

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978b). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of physiology*, 283(1):53–77.

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, 11(5):2–2.

Murray, R. F. (2013). *The Statistics of Shape, Reflectance, and Lighting in Real-World Scenes*, pages 225–235. Springer London, London.

Nestares, O. and Heeger, D. J. (1997). Modeling the apparent frequency-specific suppression in simple cell responses. *Vision research*, 37(11):1535–1543.

Nguyen, A., Yosinski, J., and Clune, J. (2016). Multifaceted feature visualization: Uncovering the different types of features learned by each neuron in deep neural networks. *arXiv preprint arXiv:1602.03616*.

Olmos, A. and Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, 33:1463 – 1473.

Olshausen, B. A. (2013). Highly overcomplete sparse coding. In *IS&T/SPIE Electronic Imaging*, pages 86510S–86510S. International Society for Optics and Photonics.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609.

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vis. Res.*, 37:3311–3325.

Olshausen, B. A. and Field, D. J. (2004). What is the other 85% of v1 doing? In Sejnowski, T. and van Hemmen, L., editors, *Problems in Systems Neuroscience*. Oxford University Press. (in press).

Olshausen, B. A. and Field, D. J. (2005). How close are we to understanding v1? *Neural computation*, 17(8):1665–1699.

Pagan, M., Simoncelli, E. P., and Rust, N. C. (2016). Neural quadratic discriminant analysis: Nonlinear decoding with v1-like computation. *Neural Computation*.

Potetz, B. and Lee, T. S. (2003). Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *JOSA A*, 20(7):1292–1303.

Prenger, R., Wu, M. C.-K., David, S. V., and Gallant, J. L. (2004). Nonlinear v1 responses to natural scenes revealed by neural network analysis. *Neural Networks*, 17(5):663–679.

Reid, R. C., Soodak, R., and Shapley, R. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology*, 66(2):505–529.

Robson, J. (1975). Receptive fields: Neural representation of the spatial and intensive attributes of the visual image. *Handbook of perception*, 5:81–116.

Rose, D. (1977). Responses of single units in cat visual cortex to moving bars of light as a function of bar length. *The Journal of physiology*, 271(1):1–23.

Rust, N. C. and Movshon, J. A. (2005). In praise of artifice. *Nature neuroscience*, 8(12):1647–1650.

Rust, N. C., Schwartz, O., Movshon, J. A., and Simoncelli, E. (2004). Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey v1. *Neurocomputing*, 58:793–799.

Sanes, J. R. and Masland, R. H. (2015). The types of retinal ganglion cells: current status and implications for neuronal classification. *Annual review of neuroscience*, 38:221–246.

Schumer, R. A. and Movshon, J. A. (1984). Length summation in simple cells of cat striate cortex. *Vision research*, 24(6):565–571.

Schwartz, O., Chichilnisky, E., and Simoncelli, E. P. (2002). Characterizing neural gain control using spike-triggered covariance. In *Advances in neural information processing systems*, pages 269–276.

Schwartz, O. and Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4:819–825.

Sclar, G. and Freeman, R. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Experimental Brain Research*, 46(3):457–461.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.

Shapley, R. and Victor, J. (1978). The effect of contrast on the transfer properties of cat retinal ganglion cells. *The Journal of physiology*, 285(1):275–298.

Shashua, A. and Ullman, S. (1990). Grouping contours by iterated pairing networks. In *Advances in neural information processing systems*, pages 335–341.

Skottun, B. C., De Valois, R. L., Grosof, D. H., Movshon, J. A., Albrecht, D. G., and Bonds, A. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision research*, 31(7):1078–1086.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9.

Tadmor, Y. and Tolhurst, D. (1989). The effect of threshold on the relationship between the receptive-field profile and the spatial-frequency tuning cure in simple cells of the cat's striate cortex. *Visual Neuroscience*, 3(05):445–454.

Tolhurst, D. and Dean, A. (1987). Spatial summation by simple cells in the striate cortex of the cat. *Experimental Brain Research*, 66(3):607–620.

Tolhurst, D. and Dean, A. (1991). Evaluation of a linear model of directional selectivity in simple cells of the cat's striate cortex. *Visual neuroscience*, 6(05):421–428.

Tolhurst, D. and Heeger, D. (1997). Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Visual Neuroscience*, 14(02):293–309.

Touryan, J., Felsen, G., and Dan, Y. (2005). Spatial structure of complex cell receptive fields measured with natural images. *Neuron*, 45(5):781–791.

van Hateren, J. H. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc.R.Soc.Lond. B*, 265:359–366.

Vecera, S. P., Vogel, E. K., and Woodman, G. F. (2002). Lower region: a new cue for figure-ground assignment. *Journal of Experimental Psychology: General*, 131(2):194–205.

Vilankar, K. P. and Field, D. J. (2017). Selectivity, hyper-selectivity and the tuning of v1 neurons. *Journal of Vision*.

Vilankar, K. P., Golden, J. R., Chandler, D. M., and Field, D. J. (2014). Local edge statistics provide information regarding occlusion and nonocclusion edges in natural scenes. *Journal of Vision*, 14(9):13–13.

Vintch, B., Movshon, J. A., and Simoncelli, E. P. (2015). A convolutional subunit model for neuronal responses in macaque v1. *Journal of Neuroscience*, 35(44):14829–14841.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE.

Webster, M. A. and Miyahara, E. (1997). Contrast adaptation and the spatial structure of natural images. *J. Opt. Soc. Am. A*, 14:2355–2366.

Webster, M. A. and Mollon, J. (1997). Adaptation and the color statistics of natural images. *Vision research*, 37(23):3283–3298.

Yang, Z. and Purves, D. (2003a). Image/source statistics of surfaces in natural scenes. *Network: Computation in Neural Systems*, 14(3):371–390.

Yang, Z. and Purves, D. (2003b). A statistical explanation of visual space. *Nature Neuroscience*, 6(6):632–640.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer.

Zetzsche, C., Krieger, G., and Wegmann, B. (1999). The atoms of vision: Cartesian or polar? *JOSA A*, 16(7):1554–1565.

Zetzsche, C. and Nuding, U. (2005). Nonlinear and higher-order approaches to the encoding of natural scenes. *Network: Computation in Neural Systems*, 16(2-3):191–221.

Zhou, C. and Mel, B. W. (2008). Cue combination and color edge detection in natural scenes. *Journal of vision*, 8(4):4.

Zhu, M. and Rozell, C. J. (2013). Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS computational biology*, 9(8):e1003191.